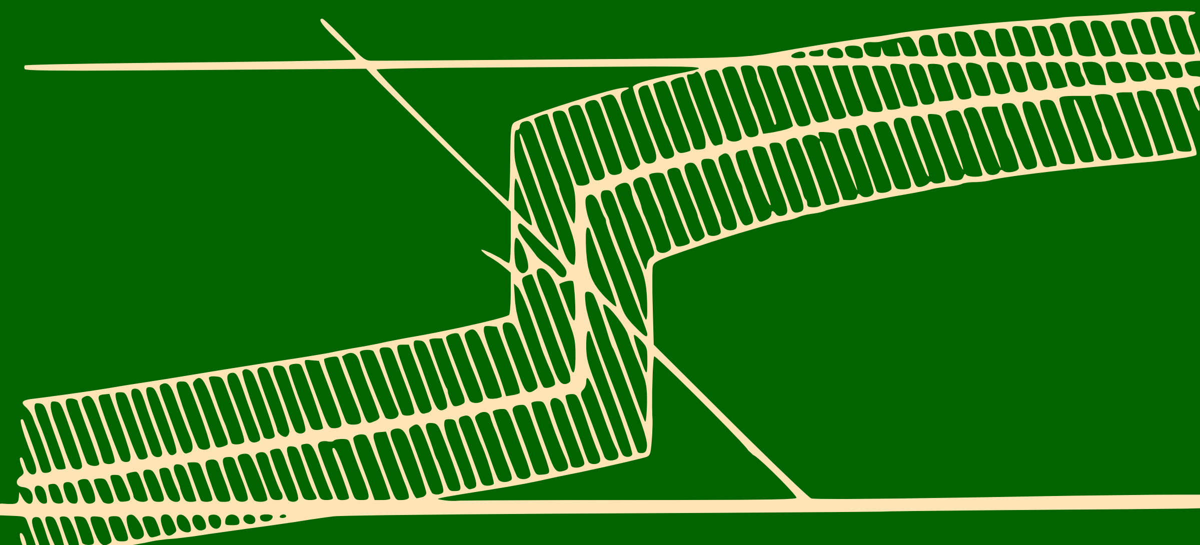
*B.V. Gnedenko*
*A.N. Kolmogorov*

# Limit Distributions For Sums of Independent Random Variables



# LIMIT DISTRIBUTIONS FOR SUMS OF INDEPENDENT RANDOM VARIABLES

ADDISON-WESLEY MATHEMATICS SERIES

ERIC REISSNER, *Consulting Editor*

---

*Bardell and Spitzbart*—COLLEGE ALGEBRA

*Dadourian*—PLANE TRIGONOMETRY

*Davis*—MODERN COLLEGE GEOMETRY

*Davis*—THE TEACHING OF MATHEMATICS

*Fuller*—ANALYTIC GEOMETRY

*Gnedenko and Kolmogorov*—LIMIT DISTRIBUTIONS FOR SUMS OF INDEPENDENT RANDOM VARIABLES

*Kaplan*—ADVANCED CALCULUS

*Kaplan*—A FIRST COURSE IN FUNCTIONS OF A COMPLEX VARIABLE

*Meserve*—FUNDAMENTAL CONCEPTS OF ALGEBRA

*Munroe*—INTRODUCTION TO MEASURE AND INTEGRATION

*Perlis*—THEORY OF MATRICES

*Stabler*—AN INTRODUCTION TO MATHEMATICAL THOUGHT

*Struik*—DIFFERENTIAL GEOMETRY

*Struik*—ELEMENTARY ANALYTIC AND PROJECTIVE GEOMETRY

*Thomas*—CALCULUS

*Thomas*—CALCULUS AND ANALYTIC GEOMETRY

*Wade*—THE ALGEBRA OF VECTORS AND MATRICES

*Wilkes, Wheeler, and Gill*—THE PREPARATION OF PROGRAMS FOR AN ELECTRONIC DIGITAL COMPUTER

# LIMIT DISTRIBUTIONS FOR SUMS OF INDEPENDENT RANDOM VARIABLES

*by*

B. V. GNEDENKO and A. N. KOLMOGOROV

*Translated from the Russian and annotated*

*by*

K. L. CHUNG

*Syracuse University*

*With an Appendix by*

J. L. DOOB

*University of Illinois*

*1954*

ADDISON-WESLEY PUBLISHING COMPANY, INC.

CAMBRIDGE 42, MASS.

# TRANSLATOR'S PREFACE

This is a translation of the Russian book ПРЕДЕЛЬНЫЕ РАСПРЕДЕЛЕНИЯ ДЛЯ СУММ НЕЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН (1949). There are various points of contact with the treatises by P. Lévy [76] and by H. Cramér [21], but much of the material in the book has been hitherto available only in periodical articles, many of which are in Russian. The systematic account presented here combines generality with simplicity, making some of the most important and difficult parts of the theory of probability easily accessible to the reader. Beyond a knowledge of the calculus on the level of, say, Hardy's *Pure Mathematics*, the book is formally self-contained. However, a certain amount of mathematical maturity, perhaps a touch of single-minded perfectionism, is needed to penetrate the depth and appreciate the classic beauty of this definitive work.

It is hoped that the English translation may serve both as a standard reference on the subject and as a text or supplementary reading for advanced courses in probability. Part of the book may also be used to suit other needs. For example, Chapters 1 and 2 may serve as the basis for any rigorous course in probability. Readers who are interested in learning the fundamental facts about stable laws and the more general infinitely divisible laws may then go on to §§ 16–18 and §§ 33–34. Those who are interested in the (weak) law of large numbers, the central limit theorem, and the analogous limit theorem leading to the Poisson law in their simpler formulations may find their needs met in § 21. Those who are interested in asymptotic expansions will need only Chapters 1, 2, 8, and 9; in particular, §§ 46–47, 49, and 51 are elementary and will be found useful for many applications.

Now a few words about the translation as compared with the original. There are two major textual changes in the English edition. The first occurs in § 32, where a mistake found in the original necessitated the deletion of several paragraphs there and thereafter. The details are explained in the second half of Appendix II. The second change occurs in §§ 46–47, where I have incorporated a substantial improvement from the 1951 Hungarian translation; see the Translator's Note to Theorem 1 of § 46.

Some minor corrections, including those of misprints, are made without mention; in a few places I have profited by the Hungarian edition which corrected some of the errors in the Russian edition. In other cases where I found fault with the Russian text, I have added a note in addition to, or instead of, changing the text. As a result, about fifty such notes are appended. These Translator's Notes are also used to supply references omitted by the authors and to add further explanatory remarks. In one

case, namely in connection with Theorem 1 of § 32, where a rather long note would be needed, I have put the added material in the first part of Appendix II.

Appendix I was written by J. L. Doob and should be of interest to the reader who may be puzzled by the measure-theoretic complications in Chapter 1.

Of the many friends who have lent me assistance of one kind or another, the following persons deserve special mention: J. L. Doob, for a variety of advice and aid; F. J. Dyson, for consultations on the Russian language; G. A. Hunt, for critically reading the manuscript; J. V. Wehausen, for helping with the Bibliography; J. Wolfowitz, for encouragement in the rather thankless job of translating. Miss Madelyn M. Keady typed the manuscript expertly and tirelessly, and my only regret is that we did not fully utilize her flawless efforts, since the formula matter was reproduced directly from the Russian edition to reduce the cost of printing. The undertaking of the translation was part of a project at Cornell University in 1952–1953, under a contract with the Air Research and Development Command, whose support is gratefully acknowledged here.

K. L. C.

CONTENTS

PREFACE . . . . . . . . 1

PART I. INTRODUCTION

CHAPTER 1. PROBABILITY DISTRIBUTIONS. RANDOM VARIABLES AND MATHEMATICAL EXPECTATIONS . . . 13

§ 1. Preliminary remarks . . . . . . . . . 13
§ 2. Measures . . . . . . . 16
§ 3. Perfect measures . . . . . . . . . . . . . . . . 18
§ 4. The Lebesgue integral . . . . . . . . . . . . . . . 19
§ 5. Mathematical foundations of the theory of probability . . . . . . 20
§ 6. Probability distributions in $R^1$ and in $R^k$ . . . . . . 22
§ 7. Independence. Composition of distributions 26
§ 8. The Stieltjes integral 29

CHAPTER 2. DISTRIBUTIONS IN $R^1$ AND THEIR CHARACTERISTIC FUNCTIONS 32

§ 9. Weak convergence of distributions 32
§ 10. Types of distributions . . . . . . . . . . . . . . . . . . 39
§ 11. The definition and the simplest properties of the characteristic function . . . . . . . . . . . . . . . . . . . . . . . . . 44
§ 12. The inversion formula and the uniqueness theorem . . . . . . . . 48
§ 13. Continuity of the correspondence between distribution and characteristic functions . . . . . . . . . . . . . . . 52
§ 14. Some special theorems about characteristic functions 55
§ 15. Moments and semi-invariants 61

CHAPTER 3. INFINITELY DIVISIBLE DISTRIBUTIONS 67

§ 16. Statement of the problem. Random functions with independent increments . . . . . . . . . . . 67
§ 17. Definition and basic properties 71
§ 18. The canonical representation . . . . . . . . . . . . . . . . . 76
§ 19. Conditions for convergence of infinitely divisible distributions 87

PART II. GENERAL LIMIT THEOREMS

CHAPTER 4. GENERAL LIMIT THEOREMS FOR SUMS OF INDEPENDENT SUMMANDS 94

§ 20. Statement of the problem. Sums of infinitely divisible summands 94
§ 21. Limit distributions with finite variances 97
§ 22. Law of large numbers 105

§ 23. Two auxiliary theorems . . . . . . . . . . . . . . . . . 109
§ 24. The general form of the limit theorems. The accompanying infinitely divisible laws . . . . . . . . . . . . . . . 112
§ 25. Necessary and sufficient conditions for convergence 116

CHAPTER 5. CONVERGENCE TO NORMAL, POISSON, AND UNITARY DISTRIBUTIONS . 125

§ 26. Conditions for convergence to normal and Poisson laws 125
§ 27. The law of large numbers 133
§ 28. Relative stability 139

CHAPTER 6. LIMIT THEOREMS FOR CUMULATIVE SUMS 145

§ 29. Distributions of the class $L$ . . . . . . . . . . . . . 145
§ 30. Canonical representation of distributions of the class $L$ . . . . . . . 149
§ 31. Conditions for convergence . . . . . . . . . . 152
§ 32. Unimodality of distributions of the class $L$ 157

PART III. IDENTICALLY DISTRIBUTED SUMMANDS

CHAPTER 7. FUNDAMENTAL LIMIT THEOREMS 162

§ 33. Statement of the problem. Stable laws . . . . . . . . . . . . . 162
§ 34. Canonical representation of stable laws 164
§ 35. Domains of attraction for stable laws 171
§ 36. Properties of stable laws . . . 182
§ 37. Domains of partial attraction 183

CHAPTER 8. IMPROVEMENT OF THEOREMS ABOUT THE CONVERGENCE TO THE NORMAL LAW 191

§ 38. Statement of the problem 191
§ 39. Two auxiliary theorems . . . . . . . . . . . . . . . 196
§ 40. Estimation of the remainder term in Lyapunov's Theorem 201
§ 41. An auxiliary theorem . . . . . . . . . . . . . . . . . . . . . 204
§ 42. Improvement of Lyapunov's Theorem for nonlattice distribution 208
§ 43. Deviation from the limit law in the case of a lattice distribution 212
§ 44. The extremal character of the Bernoulli case . . . . . . . . . . 217
§ 45. Improvement of Lyapunov's Theorem with higher moments for the continuous case . . . . . . . . . 220
§ 46. Limit theorem for densities . . . . . . . . 222
§ 47. Improvement of the limit theorem for densities 228

CHAPTER 9. LOCAL LIMIT THEOREMS FOR LATTICE DISTRIBUTIONS 231

§ 48. Statement of the problem . . . . . . . . . 231
§ 49. A local theorem for the normal limit distribution . . . . . . . 232
§ 50. A local limit theorem for non-normal stable limit distributions 235

§ 51. Improvement of the limit theorem in the case of convergence to the normal distribution 240

APPENDIX I. NOTES ON CHAPTER 1 245

APPENDIX II. NOTES ON § 32 252

BIBLIOGRAPHY 257

INDEX 262

# PREFACE

## 1

In the formal construction of a course in the theory of probability, limit theorems appear as a kind of superstructure over elementary chapters, in which all problems have finite, purely arithmetical character. In reality, however, the epistemological value of the theory of probability is revealed only by limit theorems. Moreover, without limit theorems it is impossible to understand the real content of the primary concept of all our sciences — the concept of probability. In fact, all epistemologic value of the theory of probability is based on this: that large-scale random phenomena in their collective action create strict, nonrandom regularity. The very concept of mathematical *probability* would be fruitless if it did not find its realization in the *frequency* of occurrence of events under large-scale repetition of uniform conditions (a realization which is always approximate and not wholly reliable, but that becomes, in principle, arbitrarily precise and reliable as the number of repetitions increases).

Therefore the elementary arithmetical calculations of probabilities relating to games of chance, in the works of Pascal and Fermat, can be considered only as the pre-history of the theory of probability, while its proper history began with the limit theorems of Bernoulli ([3], 1713) and de Moivre ([86], 1730). The fundamental importance of the result of de Moivre was completely revealed by Laplace ([72], 1812). To the limit theorems of Bernoulli and de Moivre-Laplace it is natural to add three more limit theorems of Poisson as the principal achievements of the theory of probability before Chebyshev. One of them generalizes the theorem of Bernoulli, another the theorem of de Moivre-Laplace, and the third leads to the so-called Poisson law of distribution. For a clear understanding of what follows it is useful to cite here somewhat modernized formulations of the five limit theorems enumerated above.

The first four deal with a sequence of *independent* events

$$\mathscr{E}_1, \mathscr{E}_2, \mathscr{E}_3, \ldots$$

We shall denote the probabilities of these events by

$$p_n = \mathsf{P}(\mathscr{E}_n),$$

and the number of actually occurring events among the first $n$ events

$$\mathscr{E}_1, \mathscr{E}_2, \ldots, \mathscr{E}_n$$

by $\mu_n$. In the first two theorems all $p_n$ have the same value $p$ ($p \neq 0$, $p \neq 1$).

1. BERNOULLI'S THEOREM. For every $\epsilon > 0$

$$\mathsf{P}\left(\left|\frac{\mu_n}{n} - p\right| > \varepsilon\right) \to 0$$

as $n \to \infty$.

2. LAPLACE'S THEOREM.

$$\mathsf{P}\left\{z_1 < \frac{\mu_n - np}{\sqrt{np(1-p)}} < z_2\right\} \to \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{z^2}{2}} dz$$

as $n \to \infty$ uniformly with respect to $z_1$ and $z_2$.

In the next two theorems $p_n$ may depend on $n$, but subject to the condition that the series

$$\sum p_n(1 - p_n)$$

diverges. We set

$$p_1 + p_2 + \ldots + p_n = A_n,$$
$$p_1(1 - p_1) + p_2(1 - p_2) + \ldots + p_n(1 - p_n) = B_n^2.$$

3. LAW OF LARGE NUMBERS IN POISSON'S FORM. For every $\epsilon > 0$

$$\mathsf{P}\left(\left|\frac{\mu_n}{n} - \frac{A_n}{n}\right| > \varepsilon\right) \to 0$$

as $n \to \infty$.

4. CENTRAL LIMIT THEOREM IN POISSON'S FORM.

$$\mathsf{P}\left\{z_1 < \frac{\mu_n - A_n}{B_n} < z_2\right\} \to \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{z^2}{2}} dz$$

as $n \to \infty$ uniformly with respect to $z_1$ and $z_2$.

The fifth of the theorems we are interested in deals with a scheme of events

$$\begin{array}{l}
\mathscr{E}_{11}, \\
\mathscr{E}_{21}, \mathscr{E}_{22}, \\
\mathscr{E}_{31}, \mathscr{E}_{32}, \mathscr{E}_{33}, \\
\cdots\cdots\cdots\cdots\cdots \\
\mathscr{E}_{n1}, \mathscr{E}_{n2}, \mathscr{E}_{n3}, \ldots, \mathscr{E}_{nn}, \\
\cdots\cdots\cdots\cdots\cdots
\end{array}$$

in which the events in the same row are mutually independent and have the same probability $p_n$, depending only on the index of the row. We denote by $\mu_n$ the number of events in the $n$th row which actually occur.

5. Poisson's Limit Theorem for Rare Events. If

$$np_n \to a$$

as $n \to \infty$, then

$$\mathsf{P}(\mu_n = m) \to \frac{a^m}{m!}e^{-a}.$$

By introducing the random variables

$$\xi_{\mathcal{E}} = \begin{cases} 1 \text{ if } \mathcal{E} \text{ occurs,} \\ 0 \text{ if } \mathcal{E} \text{ does not occur,} \end{cases}$$

we can write

$$\mu_n = \xi_{\mathcal{E}_1} + \xi_{\mathcal{E}_2} + \cdots + \xi_{\mathcal{E}_n},$$

in Theorems 1, 2, 3, 4, and

$$\mu_n = \xi_{\mathcal{E}_{n1}} + \xi_{\mathcal{E}_{n2}} + \cdots + \xi_{\mathcal{E}_{nn}}.$$

in Theorem 5.

This makes it possible to include all five limit theorems enumerated above as very special cases of limit theorems concerning *sums of independent random variables.*

The idea that the normal probability distribution

$$\mathsf{P}(\zeta < z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}} dz,$$

which turned out to be the limit in Theorems 2 and 4, must also appear in a more general problem about the limit distribution of the sum of a large number of individually negligible independent summands is one of the essential ideas of the theory of errors developed by Gauss. However, in the matter of rigorous proofs Gauss did not reach results equivalent to the theorem of de Moivre-Laplace.

Effective methods for the rigorous proof of limit theorems concerning sums of arbitrarily distributed independent variables were created in the second half of the nineteenth century by Chebyshev. His classical work opened a new period of development of the entire theory of probability.

## 2

All of Chebyshev's efforts were devoted to the solution of two problems. Consider a sequence of independent random variables

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

having finite mathematical expectations

$$a_n = \mathsf{M}\xi_n$$

and finite variances

$$b_n^2 = \mathsf{D}^2\xi_n = \mathsf{M}\,(\xi_n - a_n)^2.$$

Put

$$\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n,$$
$$A_n = a_1 + a_2 + \ldots + a_n,$$
$$B_n^2 = b_1^2 + b_2^2 + \ldots + b_n^2.$$

First Problem. What additional conditions ensure the *law of large numbers: for every* $\epsilon > 0$

$$\mathsf{P}\left(\left|\frac{\zeta_n}{n} - \frac{A_n}{n}\right| > \varepsilon\right) \to 0$$

as $n \to \infty$?

Second Problem. What additional conditions ensure the *central limit theorem:*

$$\mathsf{P}\left(\frac{\zeta_n - A_n}{B_n} < z\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}}\, dz$$

as $n \to \infty$ uniformly with respect to $z$?

For application to the first problem the method developed by Chebyshev in his work ([16], 1867) requires only the condition

$$B_n = o(n),$$

This is usually called Markov's condition, since Markov first pointed out clearly the degree of generality of Chebyshev's reasoning. The law of large numbers under Markov's condition not only includes Theorems 1 and 2 of Bernoulli and Poisson, but in the great majority of applications more or less completely settles the question for sums of independent summands.

The solution of the second problem was considerably harder. For it Chebyshev created the method of moments, which is one of his most important achievements in mathematics. The solution given by Chebyshev in his paper ([17], 1887) is based on a lemma which was proved only later

by Markov ([82], 1898). Soon afterwards the second problem of Chebyshev was solved by Lyapunov under considerably more general conditions by another method ([79], 1900; [80], 1901). Subsequently Markov succeeded in proving that the method of moments is capable of giving as general a result as that obtained by Lyapunov. However, the method of Lyapunov turned out in its further development to be much simpler and more powerful in application to the entire circle of problems concerning limit theorems for sums of independent variables. This is the method of characteristic functions, which is the principal method employed in our book.

The solution given by Lyapunov satisfies all the needs of the great majority of applications. Nevertheless, we shall give instead of Lyapunov's theorem the solution of Chebyshev's second problem in the form of Theorem 4 of § 21. The condition used there, namely Lindeberg's condition that for every $\epsilon > 0$

$$\lim_{n \to \infty} \sum_{k=1}^{n} \mathsf{P}\{|\xi_k - a_k| \geqslant \varepsilon B_n\} = 0,$$

is somewhat broader than Lyapunov's condition. In its logical structure it is even simpler than Lyapunov's condition

$$\lim_{n \to \infty} \frac{C_n}{B_n^{2+\delta}} = 0,$$

where

$$C_n = c_1 + \dots + c_n,$$
$$c_k = \mathsf{M}|\xi_k - a_k|^{2+\delta}.$$

## 3

Let us turn to the simpler special case of a sequence

$$\xi_1, \xi_2, \dots, \xi_n, \dots$$

of independent *identically distributed* variables. In this case, the central limit theorem is applicable without any additional conditions other than the mere existence of the mathematical expectations

$$a_n = a$$

and variances

$$b_n = b$$

(see Theorem 4 of § 35). However, it is erroneous to conclude, even for the case of identically distributed summands, that there exist no really inter-

esting limit theorems in which the limit laws are different from the normal law.†

In order to show by an example that such an opinion is only deep-rooted prejudice, we now consider the simple, classical scheme of random *motion* on a straight line, corresponding to the game of "heads or tails":

$$\eta(0) = 0,$$

$$\eta(t+1) = \begin{cases} \eta(t) + 1 \text{ with probability } \frac{1}{2} \\ \eta(t) - 1 \text{ with probability } \frac{1}{2} \end{cases}$$

independently of what

$$\eta(1), \ \eta(2), \ \ldots, \ \eta(t)$$

are.

It is well known that this scheme is the simplest of a long series of random motion schemes which have great importance in the most varied applications of the theory of probability, very remote from games of chance.

We number in an increasing sequence all the values of $t$ for which

$$\eta(t) = 0.$$

We obtain (with probability one) an infinite sequence

$$0 = \tau_0 < \tau_1 < \tau_2 < \ldots < \tau_n < \ldots$$

The differences

$$\xi_n = \tau_n - \tau_{n-1}$$

form a sequence of independent and identically distributed random variables. Each of the variables $\xi_n$ takes only positive even values with probabilities

$$p_m = \mathsf{P}(\xi_n = 2m) = \frac{2m\,(2m-2)!}{2^{2m}\,(m!)^2}.$$

Since

$$p_m \sim \frac{1}{2\sqrt{\pi}\, m^{3/2}},$$

asymptotically as $n \to \infty$, the mathematical expectation

$$\mathsf{M}\xi_n = 2\sum_{m=1}^{\infty} m p_m$$

is infinite. Nevertheless, the sums

$$\tau_n = \xi_1 + \xi_2 + \ldots + \xi_n,$$

† *Translator's note*. The word "law" is taken to be synonymous with "distribution" in such contexts. In "the normal (or Poisson) law" often the corresponding type (see § 10) is meant.

with suitable normalization, are subject in the limit to a completely determined law of distribution:

$$\lim_{n\to\infty} \mathsf{P}\left(\frac{2\tau_n}{\pi n^2} < z\right) = \begin{cases} 0 & \text{for} \quad z \leqslant 0, \\ \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^z e^{-\frac{1}{2z}} z^{-\frac{3}{2}}\, dz & \text{for } z > 0 \end{cases}$$

(see in this connection Theorem 5 of § 35 and the end of § 34).

The reader should turn his attention to $n^2$ in the denominator of the expression

$$\frac{2\tau_n}{\pi n^2}.$$

In the case of the sum $\zeta_n$ of identically distributed independent variables with finite variances the denominator of the expression

$$\frac{\zeta_n - A_n}{B_n}$$

in the central limit theorem would have the order $\sqrt{n}$. Comparison of these two special cases compels us to pose this general problem: Under what conditions on identically distributed independent variables

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

can a limit relation

$$\mathsf{P}\left\{\frac{\zeta_n - A_n}{B_n} < z\right\} \to V(z)$$

hold, where $A_n$ and $B_n$ are constants, and what kind of limit laws $V(z)$ can appear?

The question about the class of limit laws which can possibly appear in the situation indicated above was completely settled by A. Ya. Khintchine. It turned out that up to linear transformations this class consists only of the normal law, occupying a special position; the unitary law

$$\varepsilon(x) = \begin{cases} 0 & \text{for} \quad x \leqslant 0, \\ 1 & \text{for} \quad x > 0; \end{cases}$$

and a family of distribution laws with infinite variances, depending on two parameters ($\alpha$ and $\beta$ in the notations of Ch. 7). All these distribution laws, called "stable" because of circumstances which are explained in § 33, deserve the most serious attention. It is probable that the scope of applied problems in which they play an essential role will become in due course rather wide.

## 4

Poisson's limit theorem for rare events should long ago have suggested that even in the case of finite variances there can exist interesting and useful limit theorems concerning sums of independent variables and leading to distribution laws essentially different from the normal. To obtain them in a systematic way, it is natural to turn to the scheme of a double sequence of random variables

$$(\xi_{n1}, \xi_{n2}, \ldots, \xi_{nm_n}), \quad n = 1, 2, 3, \ldots,$$

where the random variables of the same row are independent, and to consider the sums

$$\zeta_n = \xi_{n1} + \xi_{n2} + \ldots + \xi_{nm_n}.$$

The simplest and most important case is that in which all variables $\xi_{nk}$ in the same row are identically distributed. The problem consists as before in classifying the conditions under which a limit relation

$$\mathsf{P}\left\{\frac{\zeta_n - A_n}{B_n} < z\right\} \to V(z)$$

can hold and what kind of laws $V(z)$ can appear. Here, of course, it is natural to consider only the case where

$$m_n \to \infty.$$

It is curious that if all random variables $\xi_{nk}$ can take only two values $x'$ and $x''$ independent of the indices $n$ and $k$, then the only possible limit laws (up to a linear transformation) will be the normal law, the improper law $\epsilon(x)$ and the family of Poisson laws with one parameter $a$ (see Kozulyaev [70]).

The class of possible limit laws in such a formulation of the problem coincides with the class of infinitely divisible laws, to which Chapter 3 is devoted. Naturally, it contains all the stable laws and Poisson's law. The corresponding limit theorems are proved in Chapter 4. Here we only mention that for a better understanding of their intuitive meaning it may be useful for the reader to become acquainted with a special case treated in the book of A. Ya. Khintchine [53] under the name of "generalized limit theorem of Poisson." This elementary limit theorem leads only to those infinitely divisible laws with characteristic functions of the form

$$f(t) = \exp\left\{c \int (e^{iut} - 1)\, dF(u)\right\}$$

(see § 16). Distribution laws of this type have as many finite moments as does their generating distribution $F(u)$. It is possible to indicate many physical and technical problems leading to them.

Among infinitely divisible laws, belonging neither to the class of stable laws nor to that of laws of the special type just mentioned, we mention also a family of distributions well known in mathematical statistics. They are given by the incomplete gamma functions

$$V(z) = \begin{cases} 0 & \text{for} \quad z \leqslant 0, \\ \dfrac{1}{\Gamma(\alpha)} \displaystyle\int_0^z z^{\alpha-1} e^{-z} dz & \text{for} \quad z > 0, \end{cases}$$

depending on the parameter $\alpha > 0$ (see Example 4, § 17). To this family belongs in particular (for $\alpha = 1$), the exponential distribution

$$V(z) = \begin{cases} 0 & \text{for} \quad z \leqslant 0, \\ 1 - e^{-z} & \text{for} \quad z > 0. \end{cases}$$

If we renounce the assumption that all the random variables in the same row have the same law of distribution, then the problem of determining all possible laws $V(z)$, in its exact formulation above, becomes meaningless. The limit law $V(z)$ can be absolutely arbitrary. This is indeed natural, since now the requirement $m_n \to \infty$ is illusory. It does not prevent, for example, that in each row one single summand $\xi_{nk}$ plays the dominating role. Meaningful results, comformable to the original lofty conception of the classical limit theorems in the theory of probability, are obtained only under the following additional requirement: for every $\epsilon > 0$ there should exist constants $a_{nk}$ such that

$$\sup_{1 \leqslant k \leqslant m_n} \mathsf{P}\{|\xi_{nk} - a_{nk}| \geqslant \varepsilon B_n\} \to 0.$$

This requirement of the "asymptotic negligibility" of the variation of each individual summand in comparison with the chosen scale $B_n$ for the sum $\zeta_n$ is quite natural. In § 20 it is introduced in the particular case $B_n = 1$ under the name "asymptotic constancy."

A. Ya. Khintchine proved that with this restriction the only possible limit laws in the case of arbitrarily distributed terms are the same infinitely divisible laws as in the identically distributed case (§ 24). Therefore it is quite natural that the infinitely divisible laws turn out to be the central concept throughout the first part of this book. It seems to us that the theory of these laws and the general limit theorems connected with them

will receive in time diverse applications. One of the present authors intends soon to publish in a separate article a survey of those applications of the stable and infinitely divisible laws which have already been found.

**5**

In all practical applications limit theorems are used essentially as approximate formulas for finite though sufficiently large values of $n$. In order that such an application be completely justified, the formulas should be provided with estimates of the remainder terms. If the remainder terms decrease slowly as $n \to \infty$, then it becomes necessary to introduce, for finite $n$, corrections to the limit distribution $V(z)$. The most powerful and general method of finding such corrections is to consider the various asymptotic expansions for the distribution

$$V_n(z) = \mathsf{P}\left(\frac{\zeta_n - A_n}{B_n} < z\right).$$

For the classical central limit theorem such an asymptotic expansion with terms of order

$$1, \frac{1}{\sqrt{n}}, \left(\frac{1}{\sqrt{n}}\right)^2, \left(\frac{1}{\sqrt{n}}\right)^3, \ldots$$

was indicated by Chebyshev himself without a sound basis. The recent development of his idea is traced in Chapter 8.

In Chapter 9 are discussed various other directions for the improvement of the limit theorems. Up to now there have been great achievements only in the improvement of the classical limit theorem concerning convergence to the normal distribution. The improvement of new limit theorems is given only in the direction of "local" theorems concerning convergence to the stable laws (§ 49).

The above gives a sufficiently precise outline of the scope of questions dealt with in this book. Within these limits we strive for exhaustive completeness wherever the results achieved at the present time seem to have definitive value. In those problems where now there are only results which will probably be strengthened in the near future, or where it is likely that novel formulations will combine great generality with great simplicity, we have confined ourselves to considering the simplest special cases illustrative of the nature of the problems. For this reason we consider, for example, the question of improving the limit theorems and local limit theorems only for the identically distributed case, leaving further information to periodical articles.

Generalizations to several dimensions and to sums of dependent vari-

ables are outside the scope of our book. A complete exposition of the results obtained in these directions by Markov, S. N. Bernstein, and their followers would require another volume of the same size as this, if we treated in full the limit theorems connected with Markov chains.

The striving, peculiar to the latest researches of our chosen field, for complete generality and logical perfection, and for the discovery wherever possible of conditions which are both necessary and sufficient, seems completely justified in view of its central position in the entire theory of probability. It is natural to sharpen our methods of investigation to the fullest extent on the testing ground of these classical problems and their immediate natural generalizations. Our book therefore has a *theoretical* nature: the topics selected are investigated systematically, whether or not all their developments have applied value at the moment.

We believe, however, that a great many cases of limiting behavior which seem to be introduced here only for the sake of exhausting all logical possibilities will also receive diverse applications in time. Some indications in this direction have already been given above. At any rate, there is no doubt that the arsenal of those limit theorems which should be included in future practical handbooks must be considerably expanded in comparison with classical standards. Of course, it is necessary to make some choice. For example, "normal" convergence to the non-normal stable laws (see § 35) undoubtedly must already be considered in any comprehensive text in, say, the field of statistical physics. But the consideration of "non-normal" convergence with an irregular normalization even in the case of the normal limit law (see *loc. cit.*) would only unnecessarily overburden such a practical text.

We mention also that we confine ourselves everywhere to the estimates of the *order* of the remainder terms, or in better cases to their asymptotic behavior, instead of giving estimates in the form of precise inequalities. In order to pick out this or that limit theorem as having immediate practical value, it would be necessary to fill this gap. But to trace such estimates systematically throughout our exposition would be onerous, and we preferred not to pause at all for them in this book.

## 6

It remains for us to mention that this book is based on the lectures given by us in Moscow and Lwow Universities, but in the final editing very great help was rendered us by U. V. Prohorov, who is responsible for a very great number of essential improvements in the formulations and proofs. We express our gratitude to him for this work.

# *Part I* INTRODUCTION

## CHAPTER 1

## PROBABILITY DISTRIBUTIONS. RANDOM VARIABLES AND MATHEMATICAL EXPECTATIONS

### § 1. Preliminary Remarks

In the basic chapters of this book we study exclusively the probability of events the occurrence or nonoccurrence of which is uniquely determined by the values of a finite number of real random variables

$$\xi_1, \xi_2, \ldots, \xi_n.$$

In this kind of question it is possible to get along without any more general basic concept than the probability distribution of the random point

$$\xi = (\xi_1, \xi_2, \ldots, \xi_n)$$

in the $n$-dimensional Euclidean space $R^n$. Such a distribution is given in a most logical way by means of the function

$$P_\xi(A) = P_{\xi_1 \xi_2 \cdots \xi_n}(A)$$

of the set $A \subseteq R^n$, which is understood to be the probability of the event

$$\xi \in A.$$

We impose the following restrictions:

(a) The domain of definition $\mathfrak{S}_\xi$ of the function $P_\xi(A)$ is a Borel field of subsets of $R^n$ (see the definition in § 2), containing as an element the space $R^n$ itself.

(b) The field $\mathfrak{S}_\xi$ contains all open sets of the space $R^n$.

(c) The function $P_\xi(A)$ is non-negative and countably additive.

(d) For any set $A$ of $\mathfrak{S}_\xi$ the value $P_\xi(A)$ is equal to the infimum of the values $P_\xi(G)$ for all open sets $G$ containing $A$.

(e) $P_\xi(R^n) = 1$.

The first four of these requirements express the usual properties of measures in $R^n$. The fifth requirement indicates that probability measures are always taken to be "normalized."

This concept of $n$-dimensional probability distribution is chosen as the axiomatic basis of all further considerations in, for example, H. Cramér's book [22]. However, even in the narrow frame of an exclusive interest in finite-dimensional distributions this approach is not without defect. One can be convinced of this by reading §§ 14.2 and 14.5 of Cramér's book. His axiom 3 compels us to suppose that the basic object, whose properties must be fixed by means of axioms, is a certain unnamed collection of all random variables. In § 14.5 it is even "proved" that any Borel-measurable

function of random variables is itself a random variable. But it remains vague whether we mean here those random variables to which axioms 1, 2, and 3 refer or random variables in some new sense. It would be possible, of course, to avoid these obscurities at the expense of even stronger restrictions. We could always start off with an initial probability distribution for a "basic" set of random variables

$$\xi_1, \xi_2, \ldots, \xi_n$$

and strictly distinguish from these the "generated" random variables

$$\eta = f(\xi_1, \xi_2, \ldots, \xi_n),$$

for which the distribution laws are calculated from the basic distribution $P_\xi(A)$.

A broader and more natural perspective and also the possibility of treating all random variables considered as having equal rights are opened only by a more general approach, developed, for example, in A. N. Kolmogorov's book [65]. We shall follow this system of exposition, in which all random variables considered are functions

$$\xi = \xi(u)$$

of some abstract argument $u$. §§ 2–4 contain all we need of the general theory of probability distributions on arbitrary sets. The exposition in these sections is somewhat improved over that in [65].

A basic change in comparison with [65] is the use of the new concept of "perfect" measure, which is introduced in § 3. This apparently abstract concept is introduced to bring the general theory into closer correspondence with the usual notion of distributions in concrete spaces ($R^1$, $R^n$, etc.). The reader can find in the book [69] the proofs of all the propositions formulated in § 3. To this book the reader is referred for all the demonstrations concerning the theory of measure and Lebesgue integrals.

Together with the distribution

$$P_{\xi_1\xi_2\ldots\xi_n}(A) = \mathsf{P}\{(\xi_1, \xi_2, \ldots, \xi_n) \in A\}$$

itself we often use in the study of a system of $n$ random variables $\xi_1, \xi_2, \ldots, \xi_n$ the so-called *distribution function*

$$F_{\xi_1\xi_2\ldots\xi_n}(a_1, a_2, \ldots, a_n) = \mathsf{P}\{\xi_1 < a_1, \xi_2 < a_2, \ldots, \xi_n < a_n\}.$$

This is nothing but the value of $P_{\xi_1,\xi_2,\ldots,\xi_n}(A)$ for that part of the space $R^n$ singled out by the inequalities

$$x_k < a_k;\ k = 1, 2, \ldots, n.$$

The use of such an $n$-dimensional distribution function, on the whole, may be admitted to be an anachronism, preserved from the time when the notion of a set function was not sufficiently cultivated.

We *define* the mathematical expectation of a random variable by the formula

$$\mathsf{M}\xi = \int_U \xi \, d\mathsf{P},$$

where on the right side stands the Lebesgue integral over the set $U$ of all "elementary events." From this we easily deduce other forms of writing the mathematical expectation:

$$\mathsf{M}\xi = \int_{R^1} x P_\xi (dx) = \int x \, dF_\xi (x)$$

and in case $\xi = f(\eta)$, where $\eta$ is an auxiliary random variable,

$$\mathsf{M}\xi = \int_{R^1} f(y) P_\eta (dy) = \int f(y) \, dF_\eta (y).$$

The use of the general Lebesgue integral (over an arbitrary set and with an arbitrary measure) allows us to dispense with the elementary theory of the Stieltjes integral. To us this seems a great advantage, since a complete exposition of this last theory with all the details which are necessary for a correct basis for applications to probability, is very cumbersome (and, so far as we know, has never been given). We mean the following circumstances:

(a) In order to establish the composition formula for the distribution of sums of independent terms, the definition of the Stieltjes integral current in textbooks of analysis is insufficient (see in this connection §§ 8 and 10 and supplement I to book [32]). (b) The theory of integrals with infinite limits, which are essential in the theory of probability, requires additional stipulations in the elementary approach. (c) The equation

$$\mathsf{M}\xi = \int x \, dF_\xi (x) = \int f(y) \, dF_\eta (y)$$

with $\xi = f(\eta)$, is proved in the elementary theory in a very cumbersome way. And in order that the existence of the second integral imply the existence of the first integral, it is necessary in the definition of the integral with infinite limits to impose restrictions which are essentially foreign to the elementary theory of such integrals. (d) The elementary theory of multiple Stieltjes integrals is little developed, and the questions (a), (b), and (c) are not yet fully treated in the periodical literature. By turning to the abstract theory of Lebesgue integral the necessity for all these tiresome and narrowly special investigations is eliminated. Since the distribution $P(A)$ is uniquely determined by the distribution function $F(a)$ (see § 6), by tradition we preserve the *notation*

$$\int_{R^1} f(x) P (dx) = \int f (x) \, dF (x).$$

In § 6 it is shown that under certain conditions this integral becomes the limit of Stieltjes sums. For the calculation and estimation of integrals

these sums are often useful but with our approach they are not involved in the construction of the theory of integration.

In § 7 is proved the theorem about the composition of laws of distribution when independent summands are added. The proof, based on the theorem of Fubini, is applicable without change to the sum of independent vectors, and even to more general cases.

§ 8 contains all the necessary information about integrals

$$\int_A f(x)\,\varphi(dx) \text{ and } \int_a^b f(x)\,d\Phi(x)$$

in which $\varphi(A)$ can take negative values and $\Phi(x)$ is not monotone.

## § 2. Measures

In this section we remind the reader of the basic definitions in the theory of measure. In accordance with our later needs, we confine ourselves here to finite measures.

Definition 1. The system of sets $\mathfrak{P}$ is called a *field of sets*, if (a) there exists $U \in \mathfrak{P}$ such that $A \in \mathfrak{P}$ implies $A \subseteq U$; (b) if $A \in \mathfrak{P}$ and $B \in \mathfrak{P}$, then $A \setminus B \in \mathfrak{P}$.†

It is easy to see that $U$ is uniquely determined by the field of sets. It is the "unit" of the field $U_{\mathfrak{P}}$. To every field belongs the empty set

$$N = U \setminus U.$$

From (a) and (b) it follows further that the union and intersection of any finite number of sets of the field belong to the field.

Definition 2. The field $\mathfrak{P}$ is called a *Borel field* if the union of any countable system of sets of $\mathfrak{P}$ belongs to $\mathfrak{P}$.

It is easy to prove that the intersection of a countable number of sets of a Borel field $\mathfrak{P}$ also belongs to $\mathfrak{P}$.

To avoid misunderstanding we remark that, generally speaking, the set $\{u\}$, consisting of a single element

$$u \in U_{\mathfrak{P}},$$

may not belong to $\mathfrak{P}$.

Finally, the possibility of considering only Borel fields in many questions is based on the fact that any system $\mathfrak{S}$ of subsets of a set $U$ is contained in a unique minimal Borel field with unit $U$.

This minimal field

$$\mathfrak{P} = \mathfrak{B}_U(\mathfrak{S})$$

is called the Borel closure of the system $\mathfrak{S}$ with respect to $U$.

† *Translator's note.* The symbol $A \setminus B$ denotes the set of points in one but not in both of the sets $A, B$ .

DEFINITION 3. A *measure* is a real non-negative set function $\mu(A)$, for which

($\mu 1$) The domain of definition $\mathfrak{M}_\mu$ is a Borel field of sets.

($\mu 2$) For any finite or countable number of mutually disjoint sets $A_n \in \mathfrak{M}_\mu$ we have

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n).$$

($\mu 3$) $\mu(A) = 0$ and $B \subseteq A$ imply $B \in \mathfrak{M}_\mu$.

The set

$$U_\mu = U_{\mathfrak{M}_\mu}$$

is called the *carrier* of the measure $\mu$.

In the theory of probability, where the most diverse distributions of random variables, random vectors, and random functions are obtained from the basic probability distribution $\mathbf{P}(A)$ on the set $U$ of elementary events, the following definition has fundamental importance.

DEFINITION 4. Suppose that the single-valued function

$$u' = \xi(u)$$

maps $U$ into some set $U'$. The *measure on $U'$ generated by the mapping $\xi$ from the measure $\mu$* is the set function

$$\mu'(A') = \mu_\xi^{U'}(A'),$$

for which

(1) The domain of definition $\mathfrak{M}_{\mu'}$ consists of all sets

$$A' \subseteq U'$$

for which the complete inverse image *

$$\xi^{-1}(A) \in \mathfrak{M}_\mu;$$

(2)

$$\mu'(A') = \mu(\xi^{-1}(A')).$$

It is easy to prove that the function $\mu'$ is indeed a measure, in the sense of Definition 3, for which

$$U_{\mu'} = U'.$$

If the set $U'$ is a metric (or topological) space, the mappings $\xi$ which are measurable with respect to $\mu$ (or $\mu$-measurable), i.e., the mappings for which all inverse images

$$\xi^{-1}(G)$$

* The complete inverse image $\xi^{-1}(A')$ of a set $A' \subseteq U'$, into which no element of $U_\mu$ is mapped, is taken to be the empty set $N$.

of open sets $G \subseteq U'$ belong to $\mathfrak{M}_\mu$, deserve special attention. Thus, if $\xi$ is a measurable mapping the system of sets $\mathfrak{M}_{\mu'}$ contains all open sets of the space $U'$, hence by the property ($\mu$1) of a measure it contains also all Borel sets of the space. This remark is the starting point of the considerations of the next section.

## § 3. Perfect Measures

In the study of measures in metric (or topological) spaces it is usual to confine ourselves to the consideration of those measures which besides ($\mu$1), ($\mu$2), and ($\mu$3) of our abstract measure possess also the following properties:

($\mu$4) All open sets of the space $U_\mu$ belong to $\mathfrak{M}_\mu$.
($\mu$5) The measure $\mu(A)$ of any set $A \in \mathfrak{M}_\mu$ is equal to the infimum of the measures $\mu(G)$ of open sets $G$ containing $A$.

Both these requirements are meaningless in application to measures in an abstract set $U_\mu$. At the end of the preceding section we saw, however, that the measure $\mu'$ generated by a $\mu$-measurable mapping of an abstract set $U_\mu$ into a metric space $U'$ always possesses the property ($\mu$4). To achieve complete harmony between the abstract theory of measure and the theory of measure in metric spaces it would be desirable that measurable mappings of the carrier $U_\mu$ of an abstract measure into metric spaces yield measures possessing not only the property ($\mu$4) but also the property ($\mu$5). We now see that by means of some restriction on the class of abstract measures admitted to consideration this wish is realized, at least in application to mappings in the most common and important metric spaces. The required restriction is achieved by the following *definition*.

The measure $\mu$ is called *perfect*, if the measure

$$\mu' = \mu_\xi^{R^1}$$

possesses the property ($\mu$5) whenever $\xi$ is a measurable mapping of the set $U_\mu$ into the real line $R^1$.

The reasonableness of this definition is confirmed by the following two theorems:

Theorem 1. *If $U_\mu$ is a complete metric space with a countable basis and the measure $\mu$ possesses the property ($\mu$4), then it will be perfect if and only if it possesses the property ($\mu$5).*

Theorem 2. *Any mapping $\xi$ of the carrier $U_\mu$ of a perfect measure into an arbitrary set $U'$ generates* (in accordance with Definition 4) *a perfect measure $\mu'$.*

In case the mapping $\xi$ into the metric space $U'$ is measurable, the measure $\mu'$ necessarily possesses the property ($\mu 4$); hence by comparing Theorems 1 and 2 we deduce

COROLLARY. *If the measure is perfect, then any measurable mapping $\xi$ of the set $U_\mu$ into a complete metric space with a countable basis generates in this space a measure $\mu'$ with the properties* ($\mu 4$) *and* ($\mu 5$).

Theorem 1 shows that the class of perfect measures is broad enough for applications. The corollary of Theorems 1 and 2 shows that by confining ourselves to the general theory of complete measures and their measurable mappings, we shall obtain in the most natural and simple concrete cases only measures possessing the usual properties ($\mu 4$) and ($\mu 5$).

## § 4. THE LEBESGUE INTEGRAL

$\mu$-measurable mappings $\xi$ of the set $U_\mu$ into the real line are usually called *measurable* (with respect to $\mu$) *functions*. We shall consider as known the basic properties of the *Lebesgue integral*

$$\int_A \xi(u)\,\mu(du) \tag{1}$$

of such functions over sets $A$ belonging to the system $\mathfrak{M}_\mu$ (which are usually called *$\mu$-measurable* sets). The reader can find all necessary information about the Lebesgue integral in the book [69]. Here we remind him only of some facts, not well known but important in the theory of probability. We note that in what follows *finite* Lebesgue integrals are always meant. When we say that the integral (1) *exists*, we mean that both it and the integral

$$\int_A |\xi(u)|\,\mu(du)$$

have definite finite values.

THEOREM 1. *Suppose that the mapping $\varphi$ of the set $U_\mu$ into $U'$ generates in $U'$ the measure $\mu'$; that the set $A' \subseteq U'$ belongs to $\mathfrak{M}_{\mu'}$; that the real function $\xi'(u')$ is defined on $U'$ and measurable with respect to $\mu'$; and that*

$$\xi(u) = \xi'(\varphi(u)) \text{ and } A = \varphi^{-1}(A').$$

*Then*

$$\int_A \xi(u)\,\mu(du) = \int_{A'} \xi'(u')\,\mu'(du').$$

*Moreover, both integrals exist or neither exists.*

THEOREM 2. *If* $\xi(u) \geq 0$ *on the set* $U_\mu$ *and the integral*

$$\int_{U_\mu} \xi(u)\,\mu(du)$$

*exists, then the set function*

$$\lambda(A) = \int_A \xi(u)\,\mu(du)$$

*is a measure with*

$$U_\lambda = U_\mu, \quad \mathfrak{M}_\lambda \supseteq \mathfrak{M}_\mu.$$

*Any* $\mu$*-measurable function* $\varphi(u)$ *is also* $\lambda$*-measurable and*

$$\int_A \varphi(u)\,\lambda(du) = \int_A \xi(u)\,\varphi(u)\,\mu(du).$$

*Moreover both integrals exist or neither exists.*

THEOREM 3. *If under the conditions of Theorem* 2 $\xi(u) > 0$ *everywhere on* $U_\mu$, *then for every* $A \in \mathfrak{M}_\mu = \mathfrak{M}_\lambda$

$$\mu(A) = \int_A \frac{\lambda(du)}{\xi(u)},$$

*The classes of* $\lambda$*-measurable and* $\mu$*-measurable functions* $\varphi(u)$ *coincide and*

$$\int_A \varphi(u)\,\mu(du) = \int_A \frac{\varphi(u)}{\xi(u)}\,\lambda(du).$$

*Moreover both integrals exist or neither exists.*

## § 5. MATHEMATICAL FOUNDATIONS OF THE THEORY OF PROBABILITY

After the preparation made in the preceding sections we can state very briefly the assumptions which are necessary for the development of the theory of probability.

**The probability $\mathsf{P}(A)$ is a perfect measure satisfying the additional condition of "normalization":**

$$\mathsf{P}(U_\mathsf{P}) = 1.$$

This brief formulation takes the place of all the "axiomatics" of the theory of probability.

The elements of the fundamental set

$$U = U_\mathsf{P}$$

are called in the theory of probability *elementary events*, and a set $A$ from the system

$$\mathfrak{S} = \mathfrak{M}_\mathsf{P}$$

is a *random event*. From this it follows that "elementary events" are not "random events" (even in the case, which is by no means necessary, that the set $\{u\}$ consisting of a single elementary event $u$ belongs to the system $\mathfrak{G}$).

By comparison with [65] and other expositions of the theory of probability constructed on the same principle, the approach established here contains two new restrictions: (a) The probability must satisfy the requirement ($\mu 3$) included in the definition of a measure in § 2. (b) The probability must be perfect in the sense of § 3. Both these restrictions are useful for further developments of the theory and at the same time do not narrow the domain of really interesting applications.

The *conditional probability* of a random event $A$ relative to the occurrence of the random event $B$ with $\mathbf{P}(B) > 0$ is defined by the equation

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}.$$

It is easy to prove (see [65]) that for a fixed $B$ the conditional probability possesses all the properties of the ordinary "unconditional" probability.

A $\mathbf{P}$-measurable function $\xi(u)$ of the elementary event $u$ is called a *random variable*.

By the very definition of a measurable function (see § 4) a random variable takes only real values. Only such *real* random variables are meant in the general considerations of this chapter. But the carrying over of the definition and the simplest properties of the mathematical expectation to *complex* random variables, i.e., to measurable mappings of the fundamental set $U$ into the complex plane, presents no difficulties.

*Mathematical expectations* and *conditional mathematical expectations* are defined by the formulas

$$\mathsf{M}\xi = \int_U \xi(u)\, \mathsf{P}(du), \quad \mathsf{M}(\xi|B) = \int_U \xi(u)\, \mathsf{P}(du|B).$$

In all that follows we shall denote random variables by Greek letters $\xi$, $\eta$, $\zeta$, omitting the argument $u$. Instead of $\mathbf{P}(du)$ we shall write $d\mathbf{P}$. In accordance with these conventions the definition of mathematical expectation is written

$$\mathsf{M}\xi = \int_U \xi\, d\mathsf{P}.$$

Calculations with conditional mathematical expectations are most conveniently performed on the basis of the trivial equation

$$\mathsf{M}(\xi|B) = \frac{1}{\mathsf{P}(B)} \int_B \xi\, d\mathsf{P}.$$

We remark also that in the expressions

$$\mathsf{P}(\ldots) \quad \text{or} \quad \int_{(\ldots)} \xi\, d\mathsf{P}$$

$(\cdots)$ will always denote the set of those elementary events satisfying the relations in parentheses or under the integral sign. For example,

$$\int\limits_{|\xi| < a} \xi \, d\mathsf{P}$$

denotes the integral of $\xi(u)$ over the set of those $u$ for which $|\xi(u)| < a$.

## § 6. Probability Distributions in $R^1$ and in $R^n$

Any $n$ random variables

$$\xi_1, \xi_2, \ldots, \xi_n$$

can be considered as an $n$-dimensional *random vector*

$$\xi = (\xi_1, \xi_2, \ldots, \xi_n).$$

Obviously, $n$-dimensional vectors map the fundamental set $U$ into the $n$-dimensional coordinate space $R^n$. The measure

$$P_\xi = \mathsf{P}_\xi^{R^n}$$

arising from this mapping in accordance with Definition 4 of § 2, is called the *probability distribution* of the random vector $\xi$, or the *joint distribution* of all the random variables $\xi_1, \xi_2, \ldots, \xi_n$. In an expanded form the notation is

$$P_\xi(A) = P_{\xi_1 \xi_2 \cdots \xi_n}(A).$$

It is easy to prove that every random vector maps $U$ into $R^n$ in a measurable way, i.e., the domain of definition

$$\mathfrak{S}_\xi = \mathfrak{M}_{P_\xi}$$

of the probability distribution $P_\xi$ contains all open (and consequently also all Borel) sets of the space $R^n$. In accordance with § 3 it follows from this that the measure $P_\xi$ possesses also the property ($\mu 5$). Conversely, any measure $\mu$ with

$$U_\mu = R^n,$$

possessing the properties ($\mu 4$) and ($\mu 5$) and satisfying the condition

$$\mu(R^n) = 1, \tag{*}$$

can serve as the probability distribution of an $n$-dimensional random vector. For the proof of this assertion it is sufficient to take as the set $U$ the space $R^n$ itself, to put

$$\mathsf{P} = \mu,$$

and to take as the mapping $\xi$ the identical mapping

$$\xi(u) = u.$$

Measures in $R^n$ possessing the properties ($\mu 4$) and ($\mu 5$) and satisfying the condition of normalization (*) are called *n-dimensional distributions* and are denoted preferably by the letter $P$.

Of greatest practical importance are two special classes of $n$-dimensional distributions: continuous distributions and discrete distributions.

A distribution $P$ in $R^n$ is called *continuous* if it is representable in the form

$$P(A) = \int_A p(x)\, dx, \tag{1}$$

where the sign $dx$ denotes integration with respect to the ordinary $n$-dimensional Lebesgue measure, the function $p(x)$ a Lebesgue summable function of points (vectors) $x \in R^n$. As is well known, the function $p(x)$ is determined uniquely by the distribution $P$ up to its values on a set of measure zero. It is called the *density* of the corresponding distribution. If $P = P_\xi$ is the probability distribution of the random vector $\xi$, then

$$p(x) = p_\xi(x)$$

is the *probability density* corresponding to the vector $\xi$. In an expanded form the probability density of the vector $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ is written as

$$p_\xi(x) = p_{\xi_1 \xi_2 \ldots \xi_n}(x_1, x_2, \ldots, x_n)$$

and is called the *probability density of the system of random variables* $\xi_1, \xi_2, \ldots, \xi_n$.

The probability distribution $P$ is called *discrete* if it is representable in the form

$$P(A) = \sum_{x^{(k)} \in A} p_k, \tag{2}$$

where the $x^{(k)}$ are points of $R^n$, finite or countable in number, and the $p_k$ are non-negative numbers (their sum, of course, must equal one).

If the probability distribution $P_\xi$ of a random vector $\xi$ is discrete then all the points $x$ for which

$$\mathsf{P}(\xi = x) > 0,$$

are called its *possible values*. Obviously, only these possible values play an essential role in (2).

We now turn to distributions on the real line $R^1$, i.e., to probability distributions $P_\xi$ of single random variables $\xi$. All that was said about distributions in $R^n$ for an arbitrary $n$ is applicable to this particular case $n = 1$. As we have already mentioned in § 1, in the one-dimensional case it is often expedient to single out the value of the function $P_\xi(A)$ for the

sets $A$ consisting of all the points of the real line $R^1$ lying to the left of some fixed point $a$, i.e., the (improper) intervals

$$(-\infty;\quad a).$$

Considering these values as a function of $a$, we obtain the *distribution function*

$$F_\xi(a) = P_\xi(-\infty;\quad a) = \mathsf{P}(\xi < a)$$

of the random variable $\xi$.

The proof of the following two propositions can be found in [69]:

I. * The one-dimensional distribution $P(A)$ is uniquely determined by its corresponding distribution function

$$F(a) = P(-\infty;\quad a).$$

II. In order that the function $F(a)$ may be a distribution function, it is necessary and sufficient that it be nondecreasing for all $a$, continuous to the left, and have the limiting values

$$F(-\infty) = 0,$$
$$F(+\infty) = 1.$$

Obviously, if $P$ is a continuous distribution the corresponding distribution function is represented in the form

$$F(a) = \int_{-\infty}^{a} p(x)\,dx, \tag{3}$$

where the integral must, generally speaking, be understood in the sense of Lebesgue.

In the case of a discrete distribution, given by (2), we have

$$F(a) = \sum_{x^{(k)} < a} p_k. \tag{4}$$

Obviously,

$$p_k = F(x^{(k)} + 0) - F(x^{(k)}).$$

For an arbitrary distribution $P_\xi(A)$ the analogous formula holds:

$$\mathsf{P}(\xi = x) = F_\xi(x + 0) - F_\xi(x).$$

Since the sum of probabilities of mutually exclusive events does not exceed one, the number of points at which

$$F(x + 0) - F(x) > 0,$$

i.e., the points of discontinuity of the function $F(x)$, is at most countable. If $F(x + 0) - F(x) = 1$ for some $x$, then the distribution is called *improper*. Every distribution which is not improper is called *proper*.

---

* The proof of I is based on the property ($\mu 5$) of the measure $P(A)$, i.e., on the assumption that $P(A)$ is a perfect measure.

Probability distributions are used in calculating mathematical expectations. Thus, if the random variable $\xi$ is represented in the form of a Borel measurable function

$$\xi = f(\eta_1, \eta_2, \ldots, \eta_n)$$

of other random variables $\eta_1, \eta_2, \ldots, \eta_n$, then by Theorem 1 of § 4

$$\mathsf{M}\xi = \int_U \xi \, d\mathsf{P} = \int_{R^n} f(y) P_{\eta_1\eta_2\cdots\eta_n}(dy). \tag{5}$$

If the joint distribution of the variables $\eta_1, \eta_2, \ldots, \eta_n$ is continuous, then by Theorem 2 of § 4, Eq. (5) may be written as *

$$\mathsf{M}\xi = \int_{R^n} f(y) p_{\eta_1\eta_2\cdots\eta_n}(y)\, dy. \tag{6}$$

If $P(A)$ is a one-dimensional distribution with the corresponding distribution function $F(x)$,

$$\int_a^b f(x) P(dx) = \int_a^b f(x) dF(x)$$

will denote the integral

$$\int_{[a;\, b)} f(x) P(dx)$$

over the half-open interval

$$[a;\, b) = \{ a \leqslant x < b \}$$

If $a = -\infty$, this is the interval $(-\infty < x < b)$. The integral

$$\int_a^b f(x) dF(x)$$

as function of the upper limit is always continuous to the left and the usual relation holds:

$$\int_a^b f(x)\, dF(x) + \int_b^c f(x)\, dF(x) = \int_a^c f(x)\, dF(x).$$

If the function $f(x)$ is continuous in the closed interval $[a;\, b]$, then the integral

$$I = \int_a^b f(x)\, dF(x)$$

---

* The $n$-dimensional Lebesgue measure of the whole space $R^n$ is infinite. As proved in [69], the theorems of § 4 are applicable even to the case of measures admitting infinite values.

can be calculated by means of the Stieltjes sums

$$S = \sum_{k=1}^{n} f(x_k)[F(a_k) - F(a_{k-1})],$$

where

$$a = a_0 \leqslant x_1 \leqslant a_1 \leqslant x_2 \leqslant \ldots \leqslant x_n \leqslant a_n = b.$$

As

$$\max(a_k - a_{k-1}) \to 0$$

the sums $S$ converge to the integral $I$. The integral

$$\int_{R^1} f(x)\, P(dx) = \int f(x)\, dF(x),$$

if it exists, can be obtained as

$$\lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b f(x)\, dF(x).$$

These classical methods of calculating the integrals

$$\int_{R^1} f(x) P(dx)$$

are often useful.

## § 7. Independence. Composition of Distributions

Arbitrary measurable functions

$$\xi_1, \xi_2, \ldots, \xi_n$$

of the elementary event $u$ are called independent if for any

$$A_k \in \mathfrak{S}_{\xi_k};\ k = 1, 2, \ldots, n,$$

the following equation holds:

$$\mathsf{P}\left\{\bigcap_{k=1}^{n} (\xi_k \in A_k)\right\} = \prod_{k=1}^{n} \mathsf{P}\{\xi_k \in A_k\}. \tag{1}$$

This definition is applicable to functions $\xi_k$ which have values of any nature. They may be, for example, real, complex, or vectorial functions of the argument $u$.

We shall be especially occupied with the case in which the $\xi_k$ are real random variables, i.e., measurable mappings of the set $U$ into the real line $R^1$.

THEOREM 1. *If the random variables $\xi_1, \xi_2, \ldots, \xi_n$ are independent, then their joint distribution $P_{\xi_1\xi_2\ldots\xi_n}$ is uniquely determined by the distributions*

$$P_{\xi_k};\quad k = 1, 2, \ldots, n.$$

*As a measure in*

$$R^n = R^1 \times R^1 \times \dots \times R^1$$

$P_{\xi_1\xi_2\dots\xi_n}$ *is the product of the measures* $P_{\xi_k}$ in $R^1$:

$$P_{\xi_1\xi_2\cdots\xi_n} = P_{\xi_1} \times P_{\xi_2} \times \dots \times P_{\xi_n}. \tag{2}$$

*Proof.* For any set $A \in R^n$ of the form

$$A = A_1 \times A_2 \times \dots \times A_n,$$

where

$$A_k \in \mathfrak{S}_{\xi_k};\ k = 1,\ 2, \dots, n,$$

(1) immediately yields the formula

$$P_{\xi_1\xi_2\cdots\xi_n}(A) = P_{\xi_1}(A_1)\, P_{\xi_2}(A_2) \dots P_{\xi_n}(A_n). \tag{3}$$

All parallelepipeds in $R^n$, defined by inequalities

$$a_k < x_k < b_k;\quad k = 1,\ 2,\ \dots,\ n.$$

are sets of this form.

The measure $P_{\xi_1\xi_2\dots\xi_n}(A)$ is uniquely determined throughout its domain of definition by its values for such parallelepipeds [this is true of every measure possessing the properties ($\mu$4) and ($\mu$5)]. Formula (2) is simply the expression of this construction of the measure $P_{\xi_1\xi_2\dots\xi_n}(A)$ by measures $P_{\xi_k}$ (see [69]).

If all the distributions $P_{\xi_k}$ are continuous, then (2) becomes the classical formula of the multiplication of densities:

$$p_{\xi_1\xi_2\cdots\xi_n}(x_1,\ x_2,\ \dots,\ x_n) = \prod_{k=1}^{n} p_{\xi_k}(x_k). \tag{4}$$

In accordance with the basic theme of this book we are particularly interested not in the $n$-dimensional distribution of the variables $\xi_1, \xi_2, \dots, \xi_n$, but in the one-dimensional distribution of their sum:

$$\zeta_n = \xi_1 + \xi_2 + \dots + \xi_n.$$

In the case of independence of all the summands $\xi_1, \xi_2, \dots, \xi_n$ the variable $\xi_{k+1}$ is independent of the sum

$$\zeta_k = \xi_1 + \xi_2 + \dots + \xi_k,$$

Hence the calculation of the distribution of the sum $\xi_n$ can be performed step by step by means of the following theorem about the distribution of the sum of two independent variables:

THEOREM 2. *If the random variables* $\xi$ *and* $\eta$ *are independent of each other, then the distribution of their sum*

$$\zeta = \xi + \eta$$

*is given by the formula*

$$P_\zeta(A) = \int_{R^1} P_\xi(A - y)\, P_\eta(dy). \tag{5}$$

*Here $A - y$ denotes the set of $x$ for which*

$$x + y \in A.$$

For the proof of the theorem it is sufficient to note that

$$P_\zeta(A) = \mathsf{P}(\xi + \eta \in A) = P_{\xi\eta}(B),$$

where $B$ is the set of points $(x, y)$ of the plane $R^2$ for which

$$x + y \in A.$$

By the theorem of Fubini,

$$P_{\xi\eta}(B) = \int_{R^1} P_\xi(B_y) P_\eta(dy), \tag{*}$$

where $B_y$ is the set of $x$ for which

$$(x, y) \in B.$$

In our case

$$B_y = A - y.$$

Therefore (5) follows from (*).

Since

$$(-\infty;\ z) - y = (-\infty;\ z - y),$$

for

$$F_\zeta(z) = P_\zeta(-\infty;\ z)$$

we find from (5) that

$$F_\zeta(z) = \int F_\xi(z - y)\, dF_\eta(y). \tag{6}$$

If the distributions $P_\xi$ and $P_\eta$ are continuous, then (6) becomes

$$p_\zeta(z) = \int p_\xi(z - y) p_\eta(y)\, dy.$$

Digressing from the addition of independent summands, we shall call

$$P(A) = \int_{R^1} P_1(A - y) P_2(dy),$$

$$F(z) = \int F_1(z - y)\, dF_2(y),$$

$$p(z) = \int p_1(z - y) p_2(y)\, dy$$

*compositions* of distributions, distribution functions, and densities and denote them by

$$P = P_1 * P_2,$$

$$F = F_1 * F_2,$$

$$p = p_1 * p_2.$$

This operation is also studied in analysis quite apart from the theory of probability. We shall use a mixed method, deriving the various properties of the composition of distributions sometimes purely analytically and sometimes by taking into consideration the properties of independent random variables. This is made possible by introducing for any finite set of distributions

$$P_1, P_2, \ldots, P_n$$

the distribution in $R^n$

$$P = P_1 \times P_2 \times \cdots \times P_n.$$

As already mentioned in § 5, taking the space $R^n$ to be the set $U$ of elementary events and putting

$$\mathsf{P} = P,$$

$$\xi_k(u_1, u_2, \ldots, u_n) = u_k,$$

we obtain random variables $\xi_1, \xi_2, \ldots, \xi_n$ with the joint distribution

$$P_{\xi_1\xi_2\ldots\xi_n} = P.$$

By the very definition of the distribution $P$ these random variables will be independent and will have the distributions

$$P_{\xi_k} = P_k.$$

It follows in particular that *the composition of distributions is commutative and associative.*† From the point of view of probability this is obvious. A purely analytical proof is not difficult either.

## § 8. The Stieltjes Integral

A considerable part of the theory of Lebesgue integral can be transferred to integrals of the more general form

$$\int_A f(x)\,\varphi\,(dx),$$

where $\varphi(A)$ is a countably additive set function which is real but, in contrast to a measure, is capable of taking negative values.

Countably additive set functions are usually defined on some Borel field of sets $\mathfrak{M}_\varphi$, whose unit we shall denote by $U_\varphi$ (see [69]). For any set $A \subseteq U_\varphi$, we set

$$\mu_1^*(A) = \sup_{B \subseteq A} \varphi(B), \quad \mu_2^*(A) = \sup_{B \subseteq A} [-\varphi(B)],$$

† *Translator's note.* In the original the word "distributive" was written for "associative."

where the supremum is taken over all sets $B \subseteq A$ from $\mathfrak{M}_\varphi$. These new set functions $\mu_1^*(A)$ and $\mu_2^*(A)$ are always finite, non-negative, and possess the properties of outer measures.† If they are considered only on $\mathfrak{M}_{\mu_1}$ and $\mathfrak{M}_{\mu_2}$, the Borel fields of sets measurable with respect to $\mu_1^*$ and with respect to $\mu_2^*$, then they become measures $\mu_1(A)$ and $\mu_2(A)$. We have

$$\mathfrak{M}_\varphi \subseteq \mathfrak{M}_{\mu_1} \cap \mathfrak{M}_{\mu_2}$$

and on $\mathfrak{M}_\varphi$

$$\varphi(A) = \mu_1(A) - \mu_2(A). \tag{1}$$

By definition,

$$\int_A f(x)\,\varphi(dx) = \int_A f(x)\,\mu_1(dx) - \int_A f(x)\,\mu_2(dx), \tag{2}$$

where $\mu_1$ and $\mu_2$ are the measures in the canonical decomposition of the function $\varphi$. Moreover, the integral on the left side exists (by definition) if and only if both integrals on the right side exist. For an arbitrary decomposition (1) the formula (2) is valid if both integrals on the right side exist, but the integral on the left side may exist even when the integrals on the right side do not.

In the one-dimensional case $U_\varphi$ is the real line. If the measures $\mu_1$ and $\mu_2$ satisfy the requirements ($\mu 4$) and ($\mu 5$), and if the set $A$ is the half-open interval $[a;b)$ (or the interval $(-\infty;b)$ in case $a = -\infty$), the integral (2) is written as

$$\int_a^b f(x)\,d\,\Phi(x),$$

where

$$\Phi(x) = \varphi(-\infty;\ x).$$

Now

$$\Phi(x) = \mathrm{M}_1(x) - \mathrm{M}_2(x),$$

where

$$\mathrm{M}_1(x) = \mu_1(-\infty;\ x),\ \mathrm{M}_2(x) = \mu_2(-\infty;\ x)$$

are monotone functions satisfying

$$\mathrm{M}_1(-\infty) = \mathrm{M}_2(-\infty) = 0,\ \mathrm{M}_1(+\infty) = \mu_1(R^1),\ \mathrm{M}_2(+\infty) = \mu_2(R^1),$$

Hence $\Phi$ is of bounded variation over the entire real line, is continuous to the left at every point, and has the limiting values

$$\Phi(-\infty) = 0, \quad \Phi(+\infty) = \varphi(R^1).$$

If the function $f(x)$ is continuous, then the integral (3) can be calculated by means of Stieltjes sums, as shown at the end of § 6.

---

† They are called the *positive* and *negative variations* of the function $\varphi(A)$. Their sum, $\mu_1(A) + \mu_2(A)$, is called the *total variation* of $\varphi(A)$.

In the present exposition it seems expedient to call an integral

$$\int_A f(x)\,\varphi\,(dx),$$

a *Lebesgue* integral only if $\varphi$ is a measure (i.e., a *non-negative* countably additive set function) and to reserve the name *Stieltjes* integral for the more general integral in which $\varphi$ may take negative values.

CHAPTER 2

# DISTRIBUTIONS IN $R^1$ AND THEIR CHARACTERISTIC FUNCTIONS

## § 9. Weak Convergence of Distributions

We have seen that from the general mathematical point of view a probability distribution (or as we shall simply say henceforth, a *distribution*) on the real line $R^1$ is the special case of a *measure* given on $R^1$ and possessing the properties ($\mu$1)–($\mu$5). From the class of such measures distributions are singled out by the condition of normalization

$$\mu(R^1) = 1.$$

We remind the reader of the following definition in functional analysis.

Definition. If the measures

$$\mu_1, \mu_2, \ldots, \mu_n, \ldots$$

and $\mu$ are defined in the same metric space $U$ and possess the properties ($\mu$1)–($\mu$5), then we say that the sequence $\mu_n$ *converges weakly* to $\mu$, if for any bounded continuous function $f(x)$ defined on $U$, the following relation holds:

$$\int_U f(u)\,\mu_n(du) \to \int_U f(u)\,\mu(du).$$

Applying this definition to distributions in $R^1$, we obtain the type of convergence of distributions which plays a fundamental role in the greater part of this book. Weak convergence of $\mu_n$ to $\mu$ is denoted by the symbol

$$\mu_n \Rightarrow \mu.$$

We mention at the outset the following proposition, which we need further on:

Theorem. *If $\mu_n \Rightarrow \mu$, and if the function $g(x)$ is bounded and continuous and*

$$\lambda_n(A) \equiv \int_A g(x)\,\mu_n(dx), \quad \lambda(A) \equiv \int_A g(x)\,\mu(dx),$$

*then $\lambda_n \Rightarrow \lambda$.*

For the proof we note that $\mu_n \Rightarrow \mu$ implies

$$\int_U f(x)\,g(x)\,\mu_n(dx) \to \int_U f(x)\,g(x)\,\mu(dx),$$

for every bounded and continuous function $f(x)$. By Theorem 2 of § 4, this means that

$$\int_U f(x)\lambda_n(dx) \to \int_U f(x)\lambda(dx).$$

In dealing with one-dimensional distributions we shall frequently consider also their corresponding distribution functions

$$F(x) = P(-\infty;\ x)$$

and instead of $P_n \Rightarrow P$ we shall write

$$F_n \Rightarrow F.$$

In this case a more intuitive idea of the notion of weak convergence may be obtained from the following theorem:

THEOREM 1. *For the weak convergence $F_n \Rightarrow F$ each of the following three conditions is necessary and sufficient:*

(I) *$F_n(x) \to F(x)$ at every point $x$ which is a continuity point of the distribution function $F(x)$.*

(II) *$F_n(x) \to F(x)$ on some set $C$ which is everywhere dense on the real line.*

(III) *$L(F_n,F) \to 0$, where the distance $L(G,F)$ between two distribution functions $G$ and $F$ is defined as the infimum of all $h$ such that for all $x$*

$$F(x-h) - h \leqslant G(x) \leqslant F(x+h) + h.$$

The distance $L(G,F)$ between distribution functions was introduced by P. Lévy. In Fig. 1 is drawn the strip in which the graph of $G(x)$ must be located in order to satisfy the inequality $L(G, F) \leqslant h$.

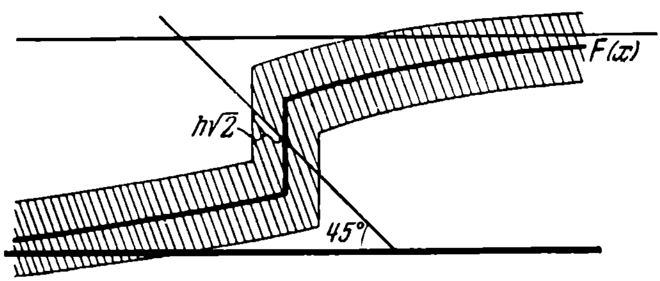



FIG. 1

It is proved in a quite elementary way that our distance satisfies the axioms of a metric:

(1) $L(F,G) = 0$ if and only if $F = G$,

(2) $L(F,G) = L(G,F)$,

(3) $L(F,H) \leq L(F,G) + L(G,H)$.

We shall carry out the proof of Theorem 1 in the following way. Denoting by (IV) the assertion of weak convergence $F_n \Rightarrow F$, we shall prove that

the following relations exist among the propositions (I), (II), (III), and (IV):

$$\begin{tikzcd}[column sep=small, row sep=small] & & (\mathrm{III}) \arrow[dr] & \\ (\mathrm{I}) \arrow[r] & (\mathrm{II}) \arrow[ur] \arrow[dr] & & (\mathrm{I}). \\ & & (\mathrm{IV}) \arrow[ur] & \end{tikzcd}$$

The assertion (I) $\rightarrow$ (II) is obvious. Hence we turn at once to the proof of the assertion (II) $\rightarrow$ (III). Take any $\epsilon > 0$ and choose $a \in C$ and $b \in C$ so that

$$F(a) \leqslant \frac{\varepsilon}{2}, \quad 1 - F(b) \leqslant \frac{\varepsilon}{2}.$$

Subdivide the closed interval $[a;b]$ by the points

$$a = a_0 < a_1 < \dots < a_s = b,$$

belonging to $C$ such that the length of each subinterval $[a_{k-1};a_k]$ is less than $\epsilon$. Choose $N$ so that for $n \geq N$ the following inequality is satisfied at all points $a_k$:

$$|F_n(a_k) - F(a_k)| \leqslant \frac{\varepsilon}{2}.$$

We now prove that for every $x$ and $n \geq N$

$$F(x - \varepsilon) - \varepsilon \leqslant F_n(x) \leqslant F(x + \varepsilon) + \varepsilon. \tag{1}$$

In the proof we consider several different cases. If

$$a_{k-1} \leqslant x \leqslant a_k,$$

then

$$F_n(x) \leqslant F_n(a_k) \leqslant F(a_k) + \frac{\varepsilon}{2} \leqslant F(x + \varepsilon) + \frac{\varepsilon}{2},$$

$$F_n(x) \geqslant F_n(a_{k-1}) \geqslant F(a_{k-1}) - \frac{\varepsilon}{2} \geqslant F(x - \varepsilon) - \frac{\varepsilon}{2}.$$

If

$$x \leqslant a_0,$$

then

$$F_n(x) \leqslant F_n(a_0) \leqslant F(a_0) + \frac{\varepsilon}{2} \leqslant \varepsilon \leqslant F(x) + \varepsilon,$$

$$F_n(x) \geqslant 0 \geqslant F(a_0) - \frac{\varepsilon}{2} \geqslant F(x) - \frac{\varepsilon}{2}.$$

If

$$x \geqslant a_s,$$

then

$$F_n(x) \leqslant 1 \leqslant F(a_s) + \frac{\varepsilon}{2} \leqslant F(x) + \frac{\varepsilon}{2},$$

$$F_n(x) \geqslant F_n(a_s) \geqslant F(a_s) - \frac{\varepsilon}{2} \geqslant 1 - \varepsilon \geqslant F(x) - \varepsilon.$$

Since $\epsilon$ is arbitrary, our assertion follows from (1), which can be rewritten as

$$L(F,\ F_n) \leqslant \varepsilon,$$

The same construction proves (II) $\rightarrow$ (IV). We denote by $M$ an upper bound of $|f(x)|$ and pick $a \in C$ and $b \in C$ so that

$$F(a) \leqslant \varepsilon, \quad 1 - F(b) \leqslant \varepsilon.$$

In the closed interval $[a;b]$ the continuous function $f(x)$ is uniformly continuous, hence there exist points $a_k$ of $C$ in this interval,

$$a = a_0 < a_1 < a_2 < \dots < a_s = b,$$

such that

$$|f(x) - f(a_k)| < \varepsilon$$

for

$$a_k \leqslant x < a_{k+1}; \quad k = 0, 1, 2, \dots, s-1.$$

Construct the auxiliary function

$$f_\varepsilon(x) = \begin{cases} f(a_k) & \text{for } a_k \leqslant x < a_{k+1}; \quad k = 0, 1, 2, \dots, s-1, \\ 0 & \text{for } x < a_0 \quad \text{or} \quad x \geqslant a_s. \end{cases}$$

Obviously, for any distribution function $G(x)$

$$\int f_\varepsilon(x)\, dG(x) = \sum_{k=0}^{s-1} f(a_k)[G(a_{k+1}) - G(a_k)].$$

Since $F_n(x) \rightarrow F(x)$ as $n \rightarrow \infty$ at the points $x = a_k$, we have

$$\int f_\varepsilon(x)\, dF_n(x) \rightarrow \int f_\varepsilon(x)\, dF(x). \tag{2}$$

At the same time, for any distribution function $G(x)$

$$\int |f(x) - f_\varepsilon(x)|\, dG(x) = \int_{-\infty}^{a} |f(x) - f_\varepsilon(x)|\, dG(x)$$
$$+ \int_a^b |f(x) - f_\varepsilon(x)|\, dG(x) + \int_b^\infty |f(x) - f_\varepsilon(x)|\, dG(x)$$
$$\leqslant MG(a) + \varepsilon[G(b) - G(a)] + M[1 - G(b)].$$

Applying this estimate to $G(x) = F(x)$ and $G(x) = F_n(x)$ and noting that $F_n(a)$ and $F_n(b)$ converge to $F(a)$ and $F(b)$, it is easy to show that for sufficiently large $n$

$$\int |f(x) - f_\varepsilon(x)|\, dF(x) \leqslant (2M + 1)\varepsilon,$$
$$\int |f(x) - f_\varepsilon(x)|\, dF_n(x) \leqslant (2M + 2)\varepsilon.$$

Together with (2) this gives

$$\left| \int f(x)\, dF_n(x) - \int f(x)\, dF(x) \right| \leqslant (4M + 4)\varepsilon$$

for sufficiently large $n$. Since $\epsilon > 0$ is arbitrary, our assertion is proved.

We shall now prove that (III) $\to$ (I). Let $x_0$ be a continuity point of $F(x)$. Then for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$|F(x) - F(x_0)| < \varepsilon$$

if

$$|x - x_0| \leqslant \delta.$$

Let

$$H = \min(\varepsilon, \delta)$$

and let $n$ be so large that $L(F_n, F) < H$. It is easy to see that

$$F_n(x_0) \geqslant F(x_0 - H) - H \geqslant F(x_0) - 2\varepsilon,$$

$$F_n(x_0) \leqslant F(x_0 + H) + H \leqslant F(x_0) + 2\varepsilon.$$

Since $\epsilon$ is arbitrary, our assertion is proved.

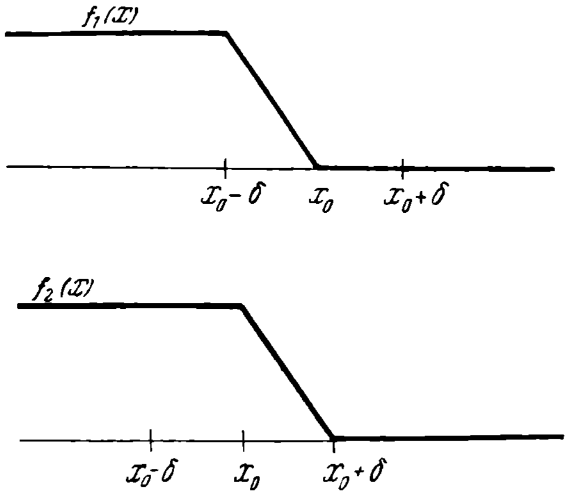



FIG. 2

Finally, we shall prove that (IV) $\to$ (I). Let $x_0$ be a continuity point of $F(x)$ and let

$$F_n \Rightarrow F.$$

Take $\delta > 0$ so that for $|x - x_0| < \delta$

$$|F(x) - F(x_0)| < \varepsilon,$$

and construct the functions (Fig. 2)

$$f_1(x) = \begin{cases} 1 & \text{for} \quad x \leqslant x_0 - \delta, \\ \dfrac{x_0 - x}{\delta} & \text{for} \quad x_0 - \delta \leqslant x \leqslant x_0, \\ 0 & \text{for} \quad x \geqslant x_0; \end{cases}$$

$$f_2(x) = \begin{cases} 1 & \text{for} \quad x \leqslant x_0, \\ 1 - \dfrac{x - x_0}{\delta} & \text{for} \quad x_0 \leqslant x \leqslant x_0 + \delta, \\ 0 & \text{for} \quad x \geqslant x_0 + \delta. \end{cases}$$

It is easy to verify that

$$\left.\begin{aligned}\int f_1(x)\,dF(x) &\geqslant \int_{-\infty}^{x_0-\delta} 1\,dF(x) = F(x_0-\delta) \geqslant F(x_0)-\varepsilon,\\ \int f_2(x)\,dF(x) &\leqslant \int_{-\infty}^{x_0+\delta} 1\,dF(x) = F(x_0+\delta) \leqslant F(x_0)+\varepsilon,\end{aligned}\right\} \tag{3}$$

$$\left.\begin{aligned}\int f_1(x)\,dF_n(x) &\leqslant \int_{-\infty}^{x_0} 1\,dF_n(x) = F_n(x_0),\\ \int f_2(x)\,dF_n(x) &\geqslant \int_{-\infty}^{x_0} 1\,dF_n(x) = F_n(x_0).\end{aligned}\right\} \tag{4}$$

For sufficiently large $n$,

$$\left.\begin{aligned}\left|\int f_1(x)\,dF_n(x) - \int f_1(x)\,dF(x)\right| &< \varepsilon,\\ \left|\int f_2(x)\,dF_n(x) - \int f_2(x)\,dF(x)\right| &< \varepsilon.\end{aligned}\right\} \tag{5}$$

It follows from (3), (4), and (5) that

$$F(x_0) - 2\varepsilon \leqslant F_n(x_0) \leqslant F(x_0) + 2\varepsilon.$$

Since $\epsilon > 0$ is arbitrary, the assertion (IV) $\rightarrow$ (I) is proved. With this the proof of Theorem 1 is also completed.

THEOREM 2. *The metric space $\mathfrak{M}^1$ of one-dimensional distributions with the distance $L(F,G)$ is complete.*

Let the sequence $F_1, F_2, \ldots, F_n, \ldots$ satisfy Cauchy's condition

$$L(F_n, F_m) \rightarrow 0 \tag{6}$$

as $n \rightarrow \infty$, $m \rightarrow \infty$. Pick an everywhere dense set

$$C = \{x_1, x_2, \ldots, x_s, \ldots\}$$

of points on the real line. Since the values of $F_n(x_s)$ are bounded, the well-known diagonal argument proves the existence of a subsequence

$$F_{n_1}(x), F_{n_2}(x), \ldots, F_{n_k}(x), \ldots,$$

which converges at every point $x = x_s$. The limit

$$v(x_s) = \lim_{k\rightarrow\infty} F_{n_k}(x_s)$$

is defined on the set $C$ and is a nondecreasing function there.

Now set

$$F(x) = \sup_{x_s < x} v(x_s).$$

The function $F(x)$ is defined everywhere on the real line, and is nondecreasing and continuous to the left. From (6) it easily follows that

$$F(-\infty)=0 \text{ and } F(+\infty)=1.$$

In fact, for any $\epsilon > 0$ there exists an $n$ such that $L(F_n,F_m) < \epsilon$ for $m \geq n$. We can find a $z$ such that $F_n(z) < \epsilon$. Then for $x_s < z - \epsilon$,

$$F_{n_k}(x_s) \leqslant F_n(z) + \varepsilon \leqslant 2\varepsilon \quad \text{for} \quad n_k \geqslant n$$

and therefore

$$v(x_s) \leqslant 2\varepsilon.$$

Since $\epsilon > 0$ is arbitrary, it follows from this that $F(-\infty) = 0$. It is similarly proved that $F(+\infty) = 1$.

It is easy to see that $F_{n_k}(x)$ converges to $F(x)$ at every continuity point of the latter function. Therefore

$$\lim_{k\to\infty} L(F_{n_k}, F) \to 0.$$

It follows from this, together with (6), that $\lim\limits_{n\to\infty} L(F_n,F) \to 0$.

THEOREM 3. *In order that the set $S$ of distributions be conditionally compact*† *in $\mathfrak{R}^1$, it is necessary and sufficient that the conditions*

$$F(x)\to 0 \quad \text{for} \quad x\to -\infty,$$
$$F(x)\to 1 \quad \text{for} \quad x\to +\infty$$

*be satisfied uniformly in $S$.*

*Proof.* Let a sequence of distribution functions $F_n(x)$ in $S$ be given. Just as in the proof of Theorem 2, we shall pick a set $C$ everywhere dense on the real line, and a subsequence

$$F_{n_1}(x),\ F_{n_2}(x),\ \ldots,\ F_{n_k}(x),\ \ldots,$$

converging to some nondecreasing, left-continuous function $F(x)$ at every continuity point of the latter. The condition of the theorem guarantees that the limit function $F(x)$ satisfies the requirement

$$F(-\infty)=0,\ F(+\infty)=1,$$

that is, $F(x)$ is a distribution function.

The definition of weak convergence at the beginning of this chapter was given for measures $\mu$ which need not satisfy the condition of normalization

$$\mu(R^1)=1,$$

characterizing a probability distribution. As was mentioned in § 8, to every measure $\mu$ in $R^1$ there corresponds a nondecreasing, left-continuous function

$$M(x)=\mu(-\infty;\ x),$$
$$M(-\infty)=0,\ M(+\infty)=\mu(R^1).$$

† *Translator's note.* The adverb "conditionally" is added. Note that a limit distribution need not belong to $S$.

In accordance with § 8, we shall write

$$\int_{R^1} f(x)\,\mu(dx) = \int f(x)\,dM(x)$$

and instead of the weak convergence $\mu_n \Rightarrow \mu$ we shall talk about the weak convergence of the corresponding functions $M_n \Rightarrow M$.

The reader can easily verify that Theorems 1, 2, 3 remain valid in the following form (see [69]).

THEOREM 1 bis. *For the weak convergence $M_n \Rightarrow M$ each of the three following conditions is necessary and sufficient:*

(I) *$M_n(x) \to M(x)$ at every point $x$ which is a continuity point of the function $M(x)$ and $M_n(+\infty) \to M(+\infty)$.*

(II) *$M_n(x) \to M(x)$ on some set $C$ everywhere dense on the real line.*†

(III) *$L(M_n, M) \to 0$, where the distance $L(M_1, M_2)$ between two functions $M_1$ and $M_2$ is defined as the infimum of all $h$ such that for all $x$*

$$M_1(x-h) - h \leqslant M_2(x) \leqslant M_1(x+h) + h.$$

THEOREM 2 bis. *The space $\mathfrak{M}^1$ of functions $M(x)$ with the distance $L(M_1, M_2)$ is complete.*

THEOREM 3 bis. *In order that the set $S$ be conditionally compact in $\mathfrak{M}^1$, it is necessary and sufficient that the limiting values $M(+\infty)$ be bounded and that the conditions*

$$M(x) \to 0 \qquad \text{for } x \to -\infty,$$
$$M(x) \to M(+\infty) \quad \text{for } x \to +\infty$$

*be satisfied uniformly on the set $S$.*

## § 10. TYPES OF DISTRIBUTIONS

It is often useful to consider, together with the random variable $\xi$, another random variable $\eta$ connected with $\xi$ by a linear relation

$$\eta = a\xi + b$$

($a > 0$ and $b$ arbitrary). Geometrically this transformation means a change of scale and of origin. It is easily seen that the distribution functions of $\xi$ and $\eta$ are connected by the equation

$$F_\xi(x) = F_\eta(ax + b).$$

As an illustration, suppose that as the solution of a certain problem we obtain the *normal distribution*

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(z-a)^2}{2\sigma^2}}\,dz,$$

† *Translator's note.* It is necessary to add the condition $M_n(+\infty) \to M(+\infty)$.

Now, for numerical computations we are not going to make a special table for this function, but will utilize the available tables of the function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}}\, dz,$$

noticing that $F(x)$ and $\Phi(x)$ are connected by the equation

$$F(x) = \Phi\left(\frac{x-a}{\sigma}\right).$$

In this connection it is natural to make use of the following concept.

DEFINITION. The distribution functions $F_1(x)$ and $F_2(x)$ *belong to the same type*, if for some constants $a > 0$ and $b$ the following equation holds:

$$F_2(x) = F_1(ax + b),$$

or, what is the same,

$$F_1(x) = F_2\left(\frac{x}{a} - \frac{b}{a}\right).$$

Since the property of belonging to the same type is symmetrical and transitive, the totality of distribution functions falls into mutually disjoint types.

It is easy to see that all normal laws of distribution form one type, the normal type: all improper distribution functions form the *improper* type. The types of distribution functions other than the improper type are called *proper*.

THEOREM 1. *If the sequence $\{F_n(x)\}$ of distribution functions converges as $n \to \infty$ to a proper distribution function $F(x)$, then for any choice of the constants $a_n > 0$ and $b_n$ the sequence $\{F_n(a_n x + b_n)\}$ can converge to a proper distribution only if this is of the same type as $F(x)$.*

*Proof.* Suppose that as $n \to \infty$, $F_n(x) \Rightarrow F(x)$ and $F_n(a_n x + b_n) \Rightarrow G(x)$, and that both $F$ and $G$ are proper. We must prove that there exist $a > 0$ and $b$ such that

$$G(x) = F(ax + b). \tag{1}$$

First of all, pick a sequence of integers $n_1 < n_2 < \cdots < n_k < \cdots$ such that the limits $\lim_{k\to\infty} a_{n_k} = a$ and $\lim_{k\to\infty} b_{n_k} = b$ $(0 \le a \le +\infty,\ -\infty \le b \le \infty)$ exist.

Henceforth we shall consider only this sequence of indices and, to simplify the notation, we assume (without loss of generality) that

$$\lim_{k\to\infty} a_k = a, \qquad \lim_{k\to\infty} b_k = b.$$

Let us prove that $0 < a < +\infty$. Suppose that $a = +\infty$. Denote by $u$ the supremum of the number $x$ for which

$$\overline{\lim_{n\to\infty}}\,(a_n x + b_n) < +\infty.$$

For $v < x < u$

$$\overline{\lim_{n\to\infty}}\,(a_n v + b_n) \leqslant \overline{\lim_{n\to\infty}}\,(v - x)\,a_n + \overline{\lim_{n\to\infty}}\,(a_n x + b_n),$$

hence by the assumptions made above we have, for every $v < u$,

$$\lim_{n\to\infty}(a_n v + b_n) = -\infty.$$

Consequently, $G(v) = 0$ for $v < u$. For $v > u$,

$$\overline{\lim_{n\to\infty}}\,(a_n v + b_n) = \infty,$$

hence $G(v) = 1$ for $v > u$.

The assumption that $a = \infty$ contradicts the fact that $G(x)$ is proper, hence it must be rejected.

It follows readily that $b$ must also be finite. In fact, the assumptions

$$\lim\,(a_n x + b_n) = +\infty, \quad \lim\,(a_n x + b_n) = -\infty$$

lead in the first case to $G(x) \equiv 1$, and in the second to $G(x) \equiv 0$.

Now suppose that $a = 0$. In this case for every $x$ and $\epsilon > 0$

$$b - \varepsilon \leqslant a_n x + b_n \leqslant b + \varepsilon$$

for sufficiently large $n$. Hence

$$F_n(b - \varepsilon) \leqslant F_n(a_n x + b_n) \leqslant F_n(b + \varepsilon),$$

and if $\epsilon$ is chosen so that the function $F(x)$ is continuous at the points $b - \epsilon$ and $b + \epsilon$, then

$$F(b - \varepsilon) \leqslant G(x) \leqslant F(b + \varepsilon).$$

Since $x$ is arbitrary, we must have

$$F(b - \varepsilon) = 0, \quad F(b + \varepsilon) = 1,$$

that is, $F(x)$ is improper, which contradicts the condition of the theorem.

Finally, let $x$ be chosen so that $F(x)$ is continuous at the point $ax + b$ and $G(x)$ is continuous at the point $x$. Then, on the one hand,

$$\lim_{n\to\infty} F_n(a_n x + b_n) = G(x),$$

and on the other hand,

$$\lim_{n\to\infty} F_n(a_n x + b_n) = F(ax + b).$$

The last equation requires clarification. Since

$$\lim_{n\to\infty}(a_n x + b_n) = ax + b,$$

for sufficiently large $n$

$$ax + b - \varepsilon \leqslant a_n x + b_n \leqslant ax + b + \varepsilon,$$

where $\epsilon > 0$ is chosen so that the function $F$ is continuous at the points $ax + b - \epsilon$ and $ax + b + \epsilon$. Hence

$$F_n(ax + b - \varepsilon) \leqslant F_n(a_n x + b_n) \leqslant F_n(ax + b + \varepsilon)$$

and in the limit as $n \to \infty$

$$F(ax + b - \varepsilon) \leqslant \varliminf_{n\to\infty} F_n(a_n x + b_n) \leqslant \varlimsup_{n\to\infty} F_n(a_n x + b_n) \leqslant F(ax + b + \varepsilon).$$

Since $ax + b$ is a continuity point of $F(x)$, and since $\epsilon$ is arbitrary, we obtain (1) from the preceding inequality. Q.E.D.

In the following we shall need the next proposition [40].

THEOREM 2. *For a sequence of distribution functions $F_n(x)$ the relations*

$$F_n(b_n x + a_n) \Rightarrow F(x), \tag{2}$$

$$F_n(\beta_n x + \alpha_n) \Rightarrow F(x), \tag{3}$$

*as $n \to \infty$, where $b_n > 0$, $\beta_n > 0$, $a_n$, $\alpha_n$ are real constants and $F(x)$ is a proper distribution function, are satisfied simultaneously if and only if*

$$\frac{\beta_n}{b_n} \to 1, \ \frac{a_n - \alpha_n}{b_n} \to 0 \tag{4}$$

*as $n \to \infty$.*

*Proof.* We shall first prove that (2) and (4) imply (3).

Let $x_1$, $x$, and $x_2$ be continuity points of the function $F(x)$, and let $x_1 < x < x_2$. Then by (4)

$$x_1 < \frac{\beta_n}{b_n} x + \frac{\alpha_n - a_n}{b_n} < x_2,$$

$$b_n x_1 + a_n < \beta_n x + \alpha_n < b_n x_2 + a_n$$

for sufficiently large $n$. Hence we conclude that

$$F_n(b_n x_1 + a_n) \leqslant F_n(\beta_n x + \alpha_n) \leqslant F_n(b_n x_2 + a_n).$$

In the limit this gives

$$F(x_1) \leqslant \varliminf_{n\to\infty} F_n(\beta_n x + \alpha_n) \leqslant \varlimsup_{n\to\infty} F_n(\beta_n x + \alpha_n) \leqslant F(x_2).$$

As $x_1 \to x$ and $x_2 \to x$, the chain of inequalities just written down becomes (3).

We now prove that, conversely, (2) and (3) imply (4). For this purpose put $B_n = \frac{\beta_n}{b_n}$, $A_n = \frac{\alpha_n - a_n}{b_n}$, $G_n(x) = F_n(b_n x + a_n)$. It is easy to verify

that in this notation the relations (2) and (3) are transformed, as $n \to \infty$, into

$$G_n(x) \to F(x),$$

$$G_n(B_n x + A_n) \to F(x).$$

As in the proof of the preceding theorem, pick a sequence $n_1 < n_2 \cdots < n_k < \cdots$ such that as $k \to \infty$

$$A_{n_k} \to A, \; B_{n_k} \to B.$$

The argument used there proves that $A$ and $B$ must be finite numbers ($B > 0$) and that the equation

$$F(x) = F(Bx + A) \tag{5}$$

must hold. Suppose that $B \neq 1$ and consider the two possible cases.

Case $B < 1$. $n$-fold application of Eq. (5) leads to the relation

$$F(x) = F(B^n x + A(1 + B + \ldots + B^{n-1})).$$

Since $n$ is arbitrary and $\lim_{n\to\infty} B^n = 0$, we conclude from this that for any $x$

$$F(x) = F\left(\frac{A}{1-B}\right).$$

For a distribution function this equation is impossible.

Case $B > 1$ reduces to the preceding, since (5) can be written as

$$F(x) = F\left(\frac{1}{B}\, x - \frac{A}{B}\right).$$

Thus we must have $B = 1$.

Now if $A \neq 0$, then (5) leads to

$$F(x) = F(x + nA),$$

where $n$ is arbitrary. Hence we find that for every $x$

$$\begin{aligned} F(-\infty) &= \lim_{An \to -\infty} F(x + nA) = F(x) \\ &= \lim_{An \to +\infty} F(x + nA) = F(+\infty). \end{aligned}$$

Since these equations are impossible for a distribution function, it follows that $A = 0$.

Thus it is proved that as $k \to \infty$

$$A_{n_k} \to 0, \quad B_{n_k} \to 1;$$

We shall prove, moreover, that

$$A_n \to 0, \quad B_n \to 1, \quad (n \to \infty). \tag{6}$$

Suppose the contrary; then there exists a number $\delta > 0$ and a sequence of indices $n'_k$ such that at least one of the inequalities

$$\lim_{k\to\infty} | B_{n'_k} - 1 | \geqslant \delta, \quad \lim_{k\to\infty} | A_{n'_k} | \geqslant \delta \tag{7}$$

holds.

Without loss of generality we may choose this sequence of indices so that

$$A_{n'_k} \to A', \quad B_{n'_k} \to B'$$

as $k \to \infty$.

From the above it is clear that we must have $A' = 0$, $B' = 1$. But these equations contradict (7). Therefore (6), and so also (4), is proved.

## § 11. The Definition and the Simplest Properties of the Characteristic Function

It is well known from elementary courses in the theory of probability that the variance

$$\mathsf{D}^2\xi = \mathsf{M}(\xi - \mathsf{M}\xi)^2.$$

is of great service in the study of sums of independent random variables. The use of variances (for example in the proof of Chebyshev's theorem) is based upon the fundamental property that the variance of a sum of independent variables $\xi_k$ is additive:

$$\mathsf{D}^2(\xi_1 + \xi_2 + \cdots + \xi_n) = \mathsf{D}^2\xi_1 + \mathsf{D}^2\xi_2 + \cdots + \mathsf{D}^2\xi_n.$$

This property of the variance of the sum of independent variables is analogous to the property of the mathematical expectation of the sum

$$\mathsf{M}(\xi_1 + \xi_2 + \cdots + \xi_n) = \mathsf{M}\xi_1 + \mathsf{M}\xi_2 + \cdots + \mathsf{M}\xi_n$$

(for mathematical expectations the requirement of independence of the summands is superfluous). It is natural to raise the general question of finding further characteristics $A\xi$ of random variables which would possess the property of *additivity*,

$$A(\xi_1 + \xi_2 + \cdots + \xi_n) = A\xi_1 + A\xi_2 + \cdots + A\xi_n,$$

for independent summands. Success in this direction would indeed be complete, should the characteristic have a definite finite value for all random variables (in contrast with the mathematical expectation and the variance, the existence of which is an additional restriction).

The answer to this question is given by introducing the concept of the *characteristic function* of the random variable (see also § 15):

$$f_\xi(t) = \mathsf{M}\, e^{it\xi} = \int e^{itx}\, dF_\xi(x). \tag{1}$$

The carrying over of the concepts of the mathematical expectation and the integral to functions with complex values, which enter here, presents no difficulties: if $w(x) = u(x) + iv(x)$, then set

$$\int_A w(x)\,\mu(dx) = \int_A u(x)\,\mu(dx) + i\int_A v(x)\,\mu(dx).$$

Since

$$|e^{it\xi}| = 1, \tag{2}$$

the function $f_\xi(t)$ is defined for all real $t$ for every distribution function $F_\xi(x)$. We shall denote a distribution function and its corresponding characteristic function by a capital letter and the corresponding small letter.

THEOREM 1. *A characteristic function is uniformly continuous on the whole line and satisfies the conditions*

$$f(0) = 1, \ |f(t)| \leqslant 1 \quad (-\infty < t < +\infty). \tag{3}$$

*Proof.* (3) follows immediately from the definition of a characteristic function (1). It remains to prove the uniform continuity of the function $f(x)$. For this purpose we introduce an inequality which will be useful also in what follows, namely, if

$$F(A) - F(-A) \geqslant 1 - \varepsilon,$$

then

$$|f(t'') - f(t')| \leqslant A\,|t'' - t'| + 2\varepsilon. \tag{4}$$

To prove this we note that for real $z'$ and $z''$ the following inequalities hold:

$$|e^{iz''} - e^{iz'}| \leqslant |z'' - z'|, \quad \text{since} \quad \left|\frac{d}{dz}e^{iz}\right| = 1,$$

$$|e^{iz''} - e^{iz'}| \leqslant 2.$$

Therefore

$$|f(t'') - f(t')| \leqslant \int_{|x| \leqslant A} + \int_{|x| > A} |e^{it''x} - e^{it'x}|\,dF(x)$$

$$\leqslant \int_{|x| \leqslant A} |it''x - it'x|\,dF(x) + 2\,[F(-A) + 1 - F(A)]$$

$$\leqslant A\,|t'' - t'| + 2\varepsilon. \qquad \text{Q.E.D.}$$

THEOREM 2. *If $\eta = a\xi + b$, where $a$ and $b$ are constants, then the characteristic functions of the random variables $\xi$ and $\eta$ are connected by the equation*

$$f_\eta(t) = f_\xi(at)\,e^{ibt},$$

*and if $a > 0$, then their distribution functions satisfy the relation*

$$F_\eta(x) = F_\xi\left(\frac{x - b}{a}\right).$$

*Proof.* In fact,

$$f_\eta(t) = \mathsf{M}e^{it\eta} = \mathsf{M}e^{it(a\xi+b)} = e^{itb}\,\mathsf{M}e^{ita\xi} = e^{itb} f_\xi(at),$$

and for $a > 0$,

$$F_\eta(x) = \mathsf{P}\{a\xi + b < x\} = \mathsf{P}\left\{\xi < \frac{x-b}{a}\right\} = F_\xi\left(\frac{x-b}{a}\right).$$

The advantages of using characteristic functions are based mainly on their next property:

THEOREM 3. *The characteristic function of the sum of two independent random variables is the product of the characteristic functions of the summands.*

*Proof.* Obviously, together with $\xi$ and $\eta$, the random variables $e^{it\xi}$ and $e^{it\eta}$ are also independent. Therefore *

$$\mathsf{M}e^{it(\xi+\eta)} = \mathsf{M}\,(e^{it\xi}\cdot e^{it\eta}) = \mathsf{M}e^{it\xi}\cdot \mathsf{M}e^{it\eta}.$$

COROLLARY 1. *If $\xi = \xi_1 + \xi_2 + \cdots + \xi_n$, and if each summand $\xi_k$ is independent of the sum of the preceding summands $\xi_1 + \xi_2 + \cdots + \xi_{k-1}$, then the characteristic function of $\xi$ is the product of the characteristic functions of the summands.*

COROLLARY 2. *The squared modulus of a characteristic function is a characteristic function.*

*Proof.* Let $\xi$ and $\eta$ be independent and have the same law of distribution with the characteristic function $f(t)$. Then, by Theorem 2,

$$f_{-\eta}(t) = f(-t) = \overline{f(t)}$$

and by Theorem 3,

$$f_{\xi-\eta}(t) = f_\xi(t)\cdot f_{-\eta}(t) = f(t)\cdot\overline{f(t)} = |f(t)|^2.$$

Q.E.D.

EXAMPLE 1. The random variable is distributed according to the normal law with mathematical expectation $a$ and variance $\sigma^2$. The characteristic function of $\xi$ is

$$\varphi(t) = \int e^{itx}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}\,dx.$$

Substituting

$$z = \frac{x-a}{\sigma} - it\sigma$$

we reduce $\varphi(t)$ to the form

$$\varphi(t) = e^{-\frac{t^2\sigma^2}{2}+iat}\frac{1}{\sqrt{2\pi}}\int_{-\infty-it\sigma}^{\infty-it\sigma} e^{-\frac{z^2}{2}}\,dz;$$

* The mathematical expectation of the product of independent variables is the product of the mathematical expectations.

Now, it is known that for every real $\alpha$,

$$\int_{-\infty-i\alpha}^{\infty-i\alpha} e^{-\frac{z^2}{2}}dz = \sqrt{2\pi},$$

consequently

$$\varphi(t) = e^{ita-\frac{1}{2}t^2\sigma^2}$$

EXAMPLE 2. The random variable $\xi$ takes only non-negative integral values and

$$\mathsf{P}\{\xi = k\} = \frac{\lambda^k}{k!}e^{-\lambda} \quad (k = 0,\ 1,\ 2,\ \ldots),$$

where $\lambda > 0$ is a constant (Poisson's law).

The characteristic function of the random variable is

$$f(t) = \mathsf{M}e^{it\xi} = \sum_{k=0}^{\infty} e^{itk}\mathsf{P}\{\xi = k\} = \sum_{k=0}^{\infty} e^{itk}\frac{\lambda^k}{k!}e^{-\lambda}$$

$$= e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^{it})^k}{k!} = e^{-\lambda}e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}.$$

Moreover it is easy to see that

$$\mathsf{M}\xi = \lambda; \qquad \mathsf{D}^2\xi = \lambda.$$

EXAMPLE 3. The random variable $\mu$ is the number of occurrences of the event $A$ in $n$ independent trials, in each of which the probability of occurrence of $A$ is $p$.

The random variable $\mu$ can be represented as the sum

$$\mu = \mu_1 + \mu_2 + \cdots + \mu_n$$

of $n$ independent random variables, each of which takes only the two values 0 and 1, with the probabilities $q = 1 - p$ and $p$ respectively. The random variable $\mu_k$ takes the value 1 if the event $A$ occurs in the $k$th trial, and the value 0 if the event $A$ does not occur in the $k$th trial.

The characteristic function of the variable $\mu_k$ is

$$f_k(t) = \mathsf{M}e^{it\mu_k} = e^{it0}q + e^{it}p = q + pe^{it}.$$

According to Theorem 3 the characteristic function of the variable $\mu$ is

$$f(t) = \prod_{k=1}^{n} f_k(t) = (q + pe^{it})^n.$$

Let us also find the characteristic function of the variable

$$\eta = (\mu - np)/\sqrt{npq}.$$

By Theorem 2 it is

$$f_\eta(t) = e^{-it\sqrt{\frac{np}{q}}}f\left(\frac{t}{\sqrt{npq}}\right) = e^{-it\sqrt{\frac{np}{q}}}\left(q + pe^{i\frac{t}{\sqrt{npq}}}\right)^n$$

$$= \left(qe^{-it\sqrt{\frac{p}{nq}}} + pe^{it\sqrt{\frac{q}{np}}}\right)^n$$

To conclude this section we remark that the definition of the characteristic function, in the form

$$f(t) = \int_{R^1} e^{itx} \varphi(dx) = \int e^{itx} d\Phi(x)$$

applies also to an arbitrary countably additive set function $\varphi(A)$ and its corresponding (see § 8) function

$$\Phi(x) = \varphi(-\infty;\ x).$$

Besides, Theorem 3 remains valid in the following form:

THEOREM 3 bis. *If*

$$\varphi = \varphi_1 * \varphi_2,\quad \Phi = \Phi_1 * \Phi_2, \tag{4}$$

*then for all t*

$$f(t) = f_1(t) f_2(t). \tag{5}$$

Here (4) stands for

$$\varphi(A) = \int_{R^1} \varphi_1(A - y)\varphi_2(dy),$$

$$\Phi(x) = \int \Phi_1(x - y)\, d\Phi_2(y).$$

The proof of this theorem can be found in texts on Fourier transforms. We can, however, easily deduce it from Theorem 3. To this end it suffices to note that every countably additive set function $\varphi(A)$ is representable in the form

$$\varphi(A) = aP(A) + bQ(A),$$

where $P$ and $Q$ are distributions and $a$ and $b$ are constants (cf. § 8).

## § 12. THE INVERSION FORMULA AND THE UNIQUENESS THEOREM

We shall now prove that the correspondence established in § 11 between one-dimensional distributions and characteristic functions is one-to-one.

THEOREM 1. *Let $f(t)$ and $F(x)$ be the characteristic function and distribution function of the random variable $\xi$. If $x_1$ and $x_2$ are continuity points of the function $F(x)$, then* *

$$F(x_2) - F(x_1) = \frac{1}{2\pi} \lim_{c \to \infty} \int_{-c}^{c} \frac{e^{-itx_1} - e^{-itx_2}}{it} f(t)\, dt. \tag{1}$$

* The equation (1) bears the name *inversion formula*.

*Proof.* For the sake of definiteness let $x_1 < x_2$. Set

$$I_c = \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-itx_1} - e^{-itx_2}}{it} f(t)\, dt.$$

Substituting here for $f(t)$ its expression in terms of $F(x)$ and changing the order of integration, we easily find

$$I_c = \frac{1}{2\pi} \int \left[ \int_{-c}^{+c} \frac{e^{it(z-x_1)} - e^{it(z-x_2)}}{it}\, dt \right] dF(z)$$

$$= \frac{1}{\pi} \int \int_0^c \left[ \frac{\sin t(z-x_1)}{t} - \frac{\sin t(z-x_2)}{t} \right] dt\, dF(z).$$

Now, for every $\alpha$ and $c$

$$\left| \frac{1}{\pi} \int_0^c \frac{\sin \alpha t}{t}\, dt \right| = \left| \frac{1}{\pi} \int_0^{\alpha c} \frac{\sin s}{s}\, ds \right| < 1, \tag{2}$$

and for $c \to \infty$

$$\frac{1}{\pi} \int_0^c \frac{\sin \alpha t}{t}\, dt \longrightarrow \begin{cases} \frac{1}{2}, & \text{if } \alpha > 0, \\ -\frac{1}{2}, & \text{if } \alpha < 0, \end{cases} \tag{3}$$

Also, this approach to the limit is uniform with respect to $\alpha$ in every domain $\alpha > \delta > 0$ (respectively $\alpha < -\delta < 0$).

Now choose $\delta$ so small that $x_1 + \delta < x_2 - \delta$ and write $I_c$ as the sum of five integrals

$$I_c = \int_{-\infty}^{x_1-\delta} + \int_{x_1-\delta}^{x_1+\delta} + \int_{x_1+\delta}^{x_2-\delta} + \int_{x_2-\delta}^{x_2+\delta} + \int_{x_2+\delta}^{\infty} \psi(c, z; x_1, x_2)\, dF(z),$$

where

$$\psi(c, z; x_1, x_2) = \frac{1}{\pi} \int_0^c \left\{ \frac{\sin t(z-x_1)}{t} - \frac{\sin t(z-x_2)}{t} \right\} dt.$$

From (3) it follows that as $c \to \infty$

$$\psi(c, z; x_1, x_2) \longrightarrow 0 \text{ for } z < x_1 - \delta \text{ and } z > x_2 + \delta$$

and

$$\psi(c, z; x_1, x_2) \longrightarrow 1 \text{ for } x_1 + \delta < z < x_2 - \delta,$$

both limits being uniform with respect to $z$. In the intervals $(x_1 - \delta, x_1 + \delta)$ and $(x_2 - \delta, x_2 + \delta)$ we know that

$$|\psi(c, z; x_1, x_2)| \leqslant 2.$$

From the relations obtained above we conclude that for every $\delta > 0$

$$\lim_{c \to \infty} I_c = F(x_2 - \delta) - F(x_1 + \delta) + R(\delta, x_1, x_2), \tag{4}$$

where

$$|R(\delta, x_1, x_2)| \leqslant 2\{F(x_1 + \delta) - F(x_1 - \delta) + F(x_2 + \delta) - F(x_2 - \delta)\}.$$

The left side of (4) does not depend on $\delta$, and the limit on the right side as $\delta$ tends to zero is $F(x_2) - F(x_1)$, by the choice of the points $x_1$ and $x_2$. Q.E.D.

THEOREM 2. *A distribution function is uniquely determined by its characteristic function.*

*Proof.* From Theorem 1 it follows immediately that at every continuity point $x$ of the function $F(x)$ the following formula applies:

$$F(x) = \frac{1}{2\pi} \lim_{y \to -\infty} \lim_{c \to \infty} \int_{-c}^{c} \frac{e^{-ity} - e^{-itx}}{it} f(t)\, dt, \tag{5}$$

where the limit in $y$ is taken over the set of points $y$ which are continuity points of $F(y)$.

*Theorems 1 and 2 remain valid if $F(x)$ is an arbitrary left-continuous function of bounded variation, subject to the condition $F(-\infty) = 0$.* In fact, such a function can be represented in the form

$$F(x) = a_1 F_1(x) + a_2 F_2(x),$$

where $F_1$ and $F_2$ are distribution functions.* For the characteristic functions we obtain in an obvious way the corresponding equation

$$f(x) = a_1 f_1(x) + a_2 f_2(x).$$

From (1) and (5), applied to $F_1$ and $F_2$ separately, we obtain immediately (1) and (5) for $F$.

We consider some examples of the application of the last theorem.

EXAMPLE 1. *If the independent random variables $\xi_1$ and $\xi_2$ are normally distributed, then their sum $\xi = \xi_1 + \xi_2$ is also normally distributed.*

In fact, if $\mathbf{M}\xi_1 = a_1$; $\mathbf{D}^2\xi_1 = \sigma_1^2$; $\mathbf{M}\xi_2 = a_2$, $\mathbf{D}^2\xi_2 = \sigma_2^2$, then the characteristic functions of the variables $\xi_1$ and $\xi_2$ are respectively

$$f_1(t) = e^{ia_1 t - \frac{1}{2}\sigma_1^2 t^2}; \quad f_2(t) = e^{ia_2 t - \frac{1}{2}\sigma_2^2 t^2}.$$

By Theorem 3 of § 11 the characteristic function $f(t)$ of the sum $\xi = \xi_1 + \xi_2$ is

$$f(t) = f_1(t) \cdot f_2(t) = e^{it(a_1 + a_2) - \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2}.$$

---

* See § 8.

This is the characteristic function of the normal law with the mathematical expectation $a = a_1 + a_2$ and the variance $\sigma^2 = \sigma_1^2 + \sigma_2^2$. On the basis of the uniqueness theorem we conclude that the distribution function of the random variable $\xi$ is normal.

It is interesting to mention that the converse proposition is also true: If the sum of two independent random variables is normally distributed then each summand is normally distributed.*

EXAMPLE 2. *The independent random variables $\xi_1$ and $\xi_2$ take only non-negative integral values, and*

$$\mathsf{P}\{\xi_1 = k\} = \frac{\lambda_1^k e^{-\lambda_1}}{k!} \text{ and } \mathsf{P}\{\xi_2 = k\} = \frac{\lambda_2^k e^{-\lambda_2}}{k!} \qquad (k \geqslant 0).$$

In Example 2 of the preceding section we found that the characteristic functions of the variables $\xi_1$ and $\xi_2$ are respectively

$$f_1(t) = e^{\lambda_1 (e^{it}-1)},$$
$$f_2(t) = e^{\lambda_2 (e^{it}-1)}.$$

The characteristic function of the sum $\xi = \xi_1 + \xi_2$ is

$$f(t) = f_1(t) \cdot f_2(t) = e^{(\lambda_1+\lambda_2)(e^{it}-1)},$$

i.e., the characteristic function of a certain Poisson law. According to the uniqueness theorem, the variable $\xi$ is distributed according to the Poisson law with the parameter $\lambda = \lambda_1 + \lambda_2$:

$$\mathsf{P}\{\xi = k\} = \frac{(\lambda_1+\lambda_2)^k e^{-(\lambda_1+\lambda_2)}}{k!} \qquad (k \geqslant 0).$$

D. A. Raikov [88] proved the converse proposition: if the sum of independent random variables is distributed according to a Poisson law then each of them is distributed according to a Poisson law.

EXAMPLE 3. *The characteristic function of a random variable $\xi$ is real if and only if the distribution function of the random variable $\xi$ is symmetrical, i.e., if for every $x$ the following equation holds:*

$$F(x) = 1 - F(-x + 0).$$

Let the distribution function be symmetrical. Then †

---

* H. Cramér [20].

† *Translator's note.* The following formula is incorrect, as is seen by taking $F(x)$ to be 0 for $x \leq 0$ and 1 for $x > 0$. The correct formula reads

$$f_\xi(t) = F(0+) - F(0-) + 2\int_{0+}^{\infty} \cos tx \, dF(x).$$

$$f_\xi(t) = \int e^{itx}\,dF(x) = \int_{-\infty}^{0} e^{itx}\,dF(x) + \int_{0}^{\infty} e^{itx}\,dF(x)$$

$$= \int_{0}^{\infty} e^{-itx}\,d(1 - F(-x)) + \int_{0}^{\infty} e^{itx}\,dF(x)$$

$$= \int_{0}^{\infty} e^{-itx}\,dF(x) + \int_{0}^{\infty} e^{itx}\,dF(x) = 2\int_{0}^{\infty} \cos tx\,dF(x).$$

To prove the converse, we consider the random variable $\eta = -\xi$. The distribution function of the variable is

$$\mathsf{P}\{\eta < x\} = \mathsf{P}\{\xi > -x\} = 1 - F(-x + 0).$$

By Theorem 2 of § 11 the characteristic functions of the variables $\xi$ and $\eta$ are connected by the relation

$$f_\eta(t) = \overline{f_\xi(t)}.$$

Since $f_\xi(t)$ is real, so also is $f_\xi(t) = \overline{f_\xi(t)}$ and thus

$$f_\eta(t) = f_\xi(t).$$

From the uniqueness theorem we conclude that the distribution functions of the variables $\xi$ and $\eta$ coincide, i.e.,

$$F(x) = 1 - F(-x + 0),$$

Q.E.D.

## § 13. Continuity of the Correspondence between Distribution and Characteristic Functions

In § 9 it was established that the totality of one-dimensional distributions with the distance $L(F,G)$ forms a complete metric space. The convergence $F_n \Rightarrow F$ in the sense

$$L(F_n, F) \to 0$$

was defined in § 9 in several equivalent ways. According to the uniqueness theorem there is a one-to-one correspondence between distribution functions $F$ and their characteristic functions,

$$f(t) = \int e^{itx}\,dF(x).$$

Therefore, putting

$$\rho(f, g) = L(F, G),$$

where $F$ and $G$ are the distribution functions corresponding to the characteristic functions $f$ and $g$, we can at once turn the totality of characteristic functions into a metric space. The convergence $f_n \Rightarrow f$ in the sense

$$\rho(f_n, f) \to 0,$$

will then obviously be equivalent to $F_n \Rightarrow F$ for the corresponding distribution functions.

However it is natural to inquire as to the meaning of such a convergence $f_n \Rightarrow f$ from the standpoint of the properties of the functions $f_n$ and $f$ themselves. The answer is given by the following theorem, which has fundamental importance for all that follows:

THEOREM 1. *If $f_n(t)$ and $f(t)$ are characteristic functions of the distributions $P_n(A)$ and $P(A)$ and if $P_n \Rightarrow P$, then*

$$f_n(t) \to f(t)$$

*as $n \to \infty$ uniformly in every bounded interval $|t| \leq T$.*

THEOREM 2. *If $f_n(t)$ is the characteristic function of the distribution $P_n(A)$ and $f_n(t)$ converges as $n \to \infty$ for all $t$ to a continuous function $f(t)$ then the distribution $P_n(A)$ converges weakly to a distribution $P(A)$ with the characteristic function $f(t)$.*

From Theorems 1 and 2 it follows that the convergence $f_n \Rightarrow f$ can be defined as uniform convergence in every finite interval. But it can also be defined as convergence at every point $t$ without any requirement of uniformity: *within the class of characteristic functions the two definitions are equivalent.**

*Proof of Theorem* 1. Let $P_n \Rightarrow P$. From the definition of weak convergence it follows that $f_n(t) \to f(t)$ for every $t$. Since $P_n \Rightarrow P$, the $P_n$ forms a conditionally compact set in the sense of weak convergence. From Theorem 3 of § 9 and (4) of § 11 we conclude that the corresponding characteristic functions $f_n(t)$ are equi-continuous. Moreover, the $f_n(t)$ are uniformly bounded. Therefore the convergence $f_n(t) \to f(t)$ as $n \to \infty$ must be uniform in every finite interval, since Arzela's theorem is applicable to $\{f_n(t)\}$ in any finite interval.†

Before proceeding to the proof of Theorem 2, let us introduce an inequality.

*Let $\tau > 0$, $X > 0$, and $\dfrac{1}{\tau X} < 1$. Then*

$$P\{-X;+X\} \geqslant \frac{\left|\dfrac{1}{2\tau}\displaystyle\int_{-\tau}^{+\tau} f(t)\,dt\right| - \dfrac{1}{\tau X}}{1-\dfrac{1}{\tau X}}. \tag{1}$$

* For an arbitrary sequence of functions $f_n(t)$ *we define the convergence $f_n(t) \Rightarrow f(t)$ to be uniform convergence in every finite interval of $t$.*

† *Translator's note.* If the functions $f_n(t)$ are equi-continuous (in every finite interval) and if $f_n(t)$ converges for every $t$ to a continuous function, then it is almost trivial to prove that the convergence is uniform in every finite interval. Neither the uniform boundedness of $f_n(t)$ nor the deeper theorem of Arzela is needed. In fact, a direct proof of Theorem 1 is very simple (*see*, e.g., [21]).

*In particular,*

$$P\left\{-\frac{2}{\tau};\ \frac{2}{\tau}\right\} \geqslant 2\left|\frac{1}{2\tau}\int_{-\tau}^{+\tau} f(t)\,dt\right| - 1. \tag{2}$$

*Proof.*

$$\left|\frac{1}{2\tau}\int_{-\tau}^{+\tau} f(t)\,dt\right| = \left|\frac{1}{2\tau}\int_{-\tau}^{+\tau}\left(\int e^{itx}\,dP\right)dt\right|$$

$$= \left|\int\left(\int_{-\tau}^{+\tau}\frac{1}{2\tau}e^{itx}\,dt\right)dP\right| \leqslant \left|\int_{|x|\leqslant X}\right| + \left|\int_{|x|>X}\frac{1}{\tau x}\sin\tau x\,dP\right|$$

$$\leqslant P\{-X;\ +X\} + \frac{1}{\tau X}(1 - P\{-X;\ +X\}) \tag{3}$$

(in the last estimates we use the inequalities

$$\left|\frac{\sin\tau x}{\tau x}\right| \leqslant 1,\quad \left|\frac{\sin\tau x}{\tau x}\right| \leqslant \frac{1}{\tau x}\right).$$

It is easy to see that (3) is equivalent to (1).

Now let the conditions of Theorem 2 be satisfied. Then for every $\epsilon > 0$ we can find a $\tau > 0$ such that

$$\left|\frac{1}{2\tau}\int_{-\tau}^{+\tau} f(t)\,dt - 1\right| < \frac{\varepsilon}{2}.$$

Consequently,

$$\left|\frac{1}{2\tau}\int_{-\tau}^{+\tau} f_n(t)\,dt - 1\right| \leqslant \left|\frac{1}{2\tau}\int_{-\tau}^{+\tau}(f_n(t) - f(t))\,dt\right| + \left|\frac{1}{2\tau}\int_{-\tau}^{+\tau} f(t)\,dt - 1\right|$$

$$\leqslant \frac{\varepsilon}{2} + \frac{1}{2\tau}\int_{-\tau}^{+\tau}|f_n(t) - f(t)|\,dt.$$

But $f_n(t) \to f(t)$ and $|f_n(t) - f(t)| \leq 2$. Therefore by Lebesgue's theorem, for fixed $\epsilon > 0$ and $\tau > 0$ and for $n \geq n(\epsilon,\tau)$,

$$\left|\frac{1}{2\tau}\int_{-\tau}^{+\tau} f_n(t)\,dt - 1\right| \leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Noting that $P_n(-X;\ +X)$ does not decrease with increasing $X$, we conclude from (2) and Theorem 3 of § 9 that the set of distributions $\{P_n\}$ is conditionally compact.

For every convergent subsequence $\{P_{n_k}\}$ the function $f_{n_k}$ converges to some continuous function $f_*(t)$, by Theorem 1.

Then $f_*(t)$ must coincide with $f(t)$. Consequently, the conditionally compact set $\{P_n\}$ has a unique limit point, which means that

$$P_n \Longrightarrow P \quad (n \to \infty).$$

As an example of the use of Theorems 1 and 2 let us prove the integral theorem of de Moivre-Laplace.

In Example 3 of § 11 we found the characteristic function $f_n(t)$ of the random variable $\eta = (\mu - np)/\sqrt{npq}$:

$$f_n(t) = \left(qe^{-it\sqrt{\frac{p}{nq}}} + pe^{it\sqrt{\frac{q}{np}}}\right)^n.$$

Using the Taylor series expansion of $e^z$, we find that

$$qe^{-it\sqrt{\frac{p}{nq}}} + pe^{it\sqrt{\frac{q}{np}}} = 1 - \frac{t^2}{2n}(1 + R_n),$$

where

$$R_n = 2\sum_{k=3}^{\infty} \frac{1}{k!}\left(\frac{it}{\sqrt{n}}\right)^{k-2} \frac{pq^k + q(-p)^k}{\sqrt{(pq)^k}}.$$

As $n \to \infty$

$$R_n \to 0,$$

uniformly in every finite interval of $t$, hence as $n \to \infty$

$$f_n(t) = \left[1 - \frac{t^2}{2n}(1 + R_n)\right]^n \Rightarrow e^{-\frac{t^2}{2}}.$$

By Theorem 2 it follows from this that for every $x$

$$\mathsf{P}\left\{\frac{\mu - np}{\sqrt{npq}} < x\right\} \Rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$$

as $n \to \infty$.

Obviously, this relation is equivalent to the usual formulation of the integral theorem of de Moivre-Laplace.

## § 14. Some Special Theorems about Characteristic Functions

In the sequel we shall need several simple properties of the characteristic function; we proceed to derive them now.

Theorem 1. *If $f(t)$ is a characteristic function then for every $t$*

$$1 - \operatorname{Re} f(2t) \leqslant 4(1 - \operatorname{Re} f(t)).$$

*Proof.* In fact

$$\operatorname{Re} f(t) = \int \cos tx \, dF(x)$$

and consequently

$$1 - \operatorname{Re} f(2t) = \int (1 - \cos 2xt) \, dF(x) = 2\int \sin^2 xt \, dF(x)$$

$$= 2\int (1 - \cos xt)(1 + \cos xt) \, dF(x) \leqslant 4\int (1 - \cos xt) \, dF(x)$$

$$= 4(1 - \operatorname{Re} f(t)).$$

We note that if $f(t)$ is real the inequality just proved reduces to

$$1 - f(2t) \leqslant 4(1 - f(t)),$$

whence for an arbitrary characteristic function $v(t)$, by Corollary 2 of Theorem 3, § 11, we obtain

$$1 - |v(2t)|^2 \leqslant 4(1 - |v(t)|^2). \tag{1}$$

THEOREM 2. *If $f(t)$ is a characteristic function and if for some sequence $t_1, t_2, \cdots$ converging to* 0,

$$|f(t_k)| = 1,$$

*then there exists a real number $\alpha$ such that*

$$f(t) = e^{i\alpha t},$$

*Hence $f(t)$ is the characteristic function of an improper distribution.*

*Proof.* Suppose the contrary, that

$$f(t) = \int e^{itx}\, dF(x),$$

where $F(x)$ is a proper distribution function. By hypothesis,

$$f(t_k) = e^{i\lambda_k}$$

where $\lambda_k$ is a real number; hence

$$1 = f(t_k)\, e^{-i\lambda_k} = \int e^{i(t_k x - \lambda_k)} dF(x).$$

From this we conclude that

$$\int (1 - \cos(t_k x - \lambda_k))\, dF(x) = 0.$$

To satisfy this equation it is obviously necessary that the function $F(x)$ be constant in every interval of $x$ in which $\cos(t_k x - \lambda_k)$ is different from one. In other words, $F(x)$ can increase only at points of the form

$$x = \frac{2\pi s + \lambda_k}{t_k}\ .$$

where $s$ is an integer. Since $F(x)$ is a proper law, there exist at least two points of increase $x_1$ and $x_2$ of $F(x)$. According to the above,

$$x_1 t_k - \lambda_k = 2\pi s_1, \quad x_2 t_k - \lambda_k = 2\pi s_2,$$

where $s_1$ and $s_2$ are distinct integers. Hence

$$|(x_1 - x_2)\, t_k| = |2\pi(s_1 - s_2)| \geqslant 2\pi.$$

By the condition of the theorem we can choose $t_k$ as small as we please, so that the inequality obtained above leads to a contradiction. Thus $F(x)$ can increase only at one point, and the theorem is proved.

*Application* 1. The condition of the theorem is certainly satisfied if $|f(t)| = 1$ in some interval $0 \leq t \leq a$ $(a > 0)$.

*Application* 2. If $f(t_k) = 1$ for some sequence $t_k$ converging to 0 then $f(t) \equiv 1$, so that

$$F(x) = \varepsilon(x) = \begin{cases} 0 & \text{for } x \leqslant 0, \\ 1 & \text{for } x > 0. \end{cases}$$

THEOREM 3. *In order that for some sequence of constants $\alpha_n$ we have*

$$F_n(x - \alpha_n) \Rightarrow \varepsilon(x) \tag{2}$$

*as $n \to \infty$, it is necessary and sufficient that as $n \to \infty$*

$$|f_n(t)| \Rightarrow 1. \tag{3}$$

(Here $F_n(x)$ denotes a distribution function and $f_n(t)$ its characteristic function.)

*Proof.* The necessity of the condition of the theorem is almost obvious. In fact, by Theorem 2 of § 11 the characteristic function of the distribution $F_n(x - \alpha_n)$ is

$$e^{-it\alpha_n} f_n(t).$$

Hence the condition

$$F_n(x - \alpha_n) \Rightarrow \varepsilon(x)$$

implies (by Theorem 1, § 13)

$$e^{-it\alpha_n} f_n(t) \Rightarrow 1,$$

and so

$$|f_n(t)| \Rightarrow 1.$$

We shall now prove the converse proposition. From (3) it follows that as $n \to \infty$

$$|f_n(t)|^2 \Rightarrow 1.$$

In other words, the characteristic function of the difference of two independent random variables $\xi_n$ and $\eta_n$, both distributed according to the law $F_n(x)$, converges to the characteristic function of the law $\epsilon(x)$ as $n \to \infty$. This means that for every $\epsilon > 0$

$$\mathsf{P}\{|\xi_n - \eta_n| \geqslant \varepsilon\} \to 0. \tag{4}$$

Pick a number $\alpha_n$ so that

$$\mathsf{P}\{\xi_n \leqslant \alpha_n\} \geqslant \frac{1}{2} \geqslant \mathsf{P}\{\xi_n < \alpha_n\}.$$

Then for every $\epsilon > 0$

$$\begin{aligned}\mathsf{P}\{\xi_n - \eta_n \geqslant \varepsilon\} &= \mathsf{P}\{(\xi_n - \alpha_n) - (\eta_n - \alpha_n) \geqslant \varepsilon\} \\ &\geqslant \mathsf{P}\{\xi_n - \alpha_n \geqslant \varepsilon, \eta_n - \alpha_n \leqslant 0\} = \mathsf{P}\{\xi_n \geqslant \alpha_n + \varepsilon\}\,\mathsf{P}\{\eta_n \leqslant \alpha_n\} \\ &\geqslant \frac{1}{2}(1 - F_n(\alpha_n + \varepsilon)). \end{aligned} \tag{5}$$

In exactly the same way,

$$\mathsf{P}\{\xi_n - \eta_n \leqslant -\varepsilon\} \geqslant \frac{1}{2} F_n(\alpha_n - \varepsilon).$$

Hence by (4) and (5), for every $\epsilon > 0$ as $n \to \infty$

$$1 - F_n(\alpha_n + \varepsilon) + F_n(\alpha_n - \varepsilon) \leqslant 2\mathsf{P}\{|\xi_n - \alpha_n| > \varepsilon\} \to 0.$$

This relation is obviously equivalent to (2).

THEOREM 4. *If for some sequence of integers* $n_1 < n_2 < \cdots < n_k < \cdots$

$$f_k^{n_k}(t) \Rightarrow f(t) \tag{6}$$

*as* $k \to \infty$, *where* $f(t)$ *is some continuous function and the* $f_k(t)$ $(k = 1, 2, \ldots)$ *are characteristic functions, then*

$$f_k(t) \Rightarrow 1 \tag{7}$$

*as* $k \to \infty$, *so that*

$$F_k(x) \Rightarrow \varepsilon(x).$$

*Proof.* Since $f(0) = 1$ and $f(t)$ is continuous, there exists $a > 0$ such that

$$|f(t)| > 0 \quad \text{for} \quad |t| \leqslant a.$$

From (6) it is clear that in this interval

$$f_k(t) \to 1.$$

From this and the inequality in Theorem 1 of this section

$$1 - \operatorname{Re} f_k(2t) \leqslant 4(1 - \operatorname{Re} f_k(t)),$$

we conclude that

$$\operatorname{Re} f_k(2t) \to 1 \qquad \text{for} \quad |t| \leqslant a,$$

and therefore

$$f_k(t) \to 1 \qquad \text{for} \quad |t| \leqslant 2a.$$

The possibility of doubling any interval in which $f_k(t) \to 1$ clearly yields the conclusion of the theorem.

We shall call a discrete distribution of a random variable a *lattice distribution* if there exist numbers $a$ and $h > 0$ such that every possible value of $\xi$ can be represented in the form $a + kh$, where $k$ runs through integral values (not necessarily all). We shall call the number $h$ a *span of the distribution.*

Many important distributions in the theory of probability belong to the class of lattice distributions (for example, the Bernoulli distribution, the Poisson distribution).

If it is impossible to represent all the possible values of $\xi$ in the form $b + kh_1$ for some $b$ and some $h_1 > h$, then we shall say that $h$ is a *maximum span of the distribution.*

The conditions for a maximum span of a distribution can be expressed in other terms. Namely, a span $h$ will be *maximum* if and only if one is the greatest common divisor of the pairwise differences of the possible values of $\xi$, divided by $h$. A little later we shall give a third condition for a

span to be maximum. Now we proceed to establish the following characteristic property of lattice distributions.

THEOREM 5. *In order that the random variable $\xi$ have a lattice distribution, it is necessary and sufficient that for some nonzero value of the argument the modulus of the characteristic function of $\xi$ be equal to one.*

*Proof.* The necessity of the condition is proved at once by calculation. Let $\xi$ have a lattice distribution and

$$p_k = \mathsf{P}\ \{\ \xi = a + kh\ \}.$$

Then the characteristic function of $\xi$ is

$$f(t) = e^{iat} \sum_{k=-\infty}^{\infty} e^{ikht}\, p_k. \tag{8}$$

The second factor is a periodic function with the period $\frac{2\pi}{h}\cdot$ Since $f(0) = 1$, it is evident that

$$\left| f\left(\frac{2\pi}{h}\right) \right| = 1.$$

Now suppose that for $t_0 \neq 0$

$$|f(t_0)| = 1.$$

In other words, we suppose that for some real $a$

$$f(t_0) = e^{it_0 a}.$$

This equation can be written out as

$$\int e^{it_0(x-a)}\, dF(x) = 1,$$

whence it follows that

$$\int \cos t_0 (x - a)\, dF(x) = 1.$$

For this equation to hold, it is necessary that the function $F(x)$ be constant everywhere with the exception of those $x$ for which

$$\cos t_0 (x - a) = 1.$$

All $x$ satisfying the last equation have the form

$$x = a + k\frac{2\pi}{t_0}, \tag{9}$$

where $k$ is an integer. Q.E.D.

From the theorem just proved we easily deduce

COROLLARY 1. *If the characteristic function $f(t)$ is such that for two incommensurable values of the argument $t_0$ and $t_1$ the equations*

$$|f(t_0)| = 1 \text{ and } |f(t_1)| = 1$$

*both hold, then*

$$|f(t)| \equiv 1.$$

*In particular, if the modulus of the characteristic function $f(t)$ is equal to one in any interval $a \leq t \leq b$, $a < b$, then it is equal to one for all values of $t$.*

Theorem 5 enables us to formulate the following important result.

COROLLARY 2. *A span $h$ of the distribution is maximum if and only if the modulus of the characteristic function of $\xi$ is less than one in the interval $0 < |t| < 2\pi/h$ and equal to one for $t = 2\pi/h$.*

*Proof.* If

$$|f(t_1)| = 1$$

for some $t_1$, $0 < |t_1| < 2\pi/h$, then according to (9) the number $2\pi/|t_1|$ would be a span of the distribution. But since by hypothesis

$$h < \frac{2\pi}{|t_1|},$$

$h$ cannot be a maximum span.

From this we conclude that for every lattice distribution and for every $\epsilon > 0$, we can find a $c > 0$ $[c = c(\epsilon)]$ such that if $h$ is a maximum span then for $\epsilon < |t| < \dfrac{2\pi}{h} - \epsilon$

$$|f(t)| < e^{-c}.$$

For later purposes we must deduce two elementary formulas. Multiply both sides of the equation (8) by $e^{-iat-irth}$ ($r$ an integer) and integrate from $-\dfrac{\pi}{h}$ to $\dfrac{\pi}{h}$. Since

$$\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{it(k-r)h}\,dt = \begin{cases} 0 & \text{for} \quad k \neq r, \\ \dfrac{2\pi}{h} & \text{for} \quad k = r, \end{cases}$$

we obtain as a result

$$p_r = \frac{h}{2\pi} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} f(t)\, e^{-iat-irth}\,dt. \tag{10}$$

This equation enables us to write the inversion formula for lattice distributions in a somewhat different form from the one we had before. Indeed, putting

$$x_1 = a + mh - \frac{1}{2}h, \qquad x_2 = a + nh + \frac{1}{2}h \quad (n \geqslant m),$$

we shall prove that

$$F(x_2) - F(x_1) = \frac{h}{2\pi} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} f(t) \frac{e^{-itx_1} - e^{-itx_2}}{2i \sin \frac{th}{2}} dt. \tag{11}$$

In fact,

$$F(x_2) - F(x_1) = \sum_{r=m}^{n} p_r = \frac{h}{2\pi} \sum_{r=m}^{n} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} f(t) e^{-ita-itrh} dt,$$

and

$$\sum_{r=m}^{n} e^{-ita-itrh} = e^{-ita} \frac{e^{-itmh} - e^{-it(n+1)h}}{1 - e^{-ith}}$$

$$= e^{-ita} \frac{e^{-ith\left(m-\frac{1}{2}\right)} - e^{-ith\left(n+\frac{1}{2}\right)}}{2 \sin \frac{th}{2}} = \frac{e^{-itx_1} - e^{-itx_2}}{2 \sin \frac{th}{2}}$$

From Corollary 2 follows:

COROLLARY 3. *Every lattice distribution, apart from the improper ones, has a unique maximum span.*

Of course, Corollary 3 can also be easily obtained in an elementary arithmetical way.

## § 15. Moments and Semi-invariants

By virtue of the uniqueness theorem of § 12 the values of the characteristic function

$$f(t) = \int e^{itx} dF(x)$$

for all $t$ determine the distribution function $F(x)$. It is therefore natural to expect that all other numerical characteristics of a one-dimensional distribution (or distribution function) can be expressed in terms of its characteristic function. For example, we shall soon see that

$$\mathsf{M}\xi = -if'_\xi(0), \tag{1}$$

$$\mathsf{D}^2\xi = -f''_\xi(0) - [f'_\xi(0)]^2. \tag{2}$$

At the beginning of this section we shall consider from this point of view the most useful numerical characteristics of one-dimensional distributions: moments and semi-invariants.

The number

$$\alpha_s = \mathsf{M}\xi^s = \int x^s \, dF_\xi(x) \tag{3}$$

is called the *moment* of order $s$. In accordance with § 4, for the existence of the moment $\alpha$, the existence of the *absolute moment* of order $s$,

$$\beta_s = \mathsf{M}|\xi|^s = \int |x|^s \, dF_\xi(x), \tag{4}$$

is necessary and sufficient. The following lemma is well known from elementary courses in the theory of probability.

LEMMA 1. *If the moment $\beta_s$ exists for the random variable $\xi$, then all moments $\beta_k$ for $k < s$ exist and*

$$\beta_1 \leqslant \beta_2^{\frac{1}{2}} \leqslant \beta_3^{\frac{1}{3}} \leqslant \dots \leqslant \beta_s^{\frac{1}{s}}. \tag{5}$$

As for the moments $\alpha_0$ and $\beta_0$ of order zero, they are always considered to exist and to be equal:

$$\alpha_0 = \beta_0 = 1. \tag{6}$$

By definition,

$$\alpha_1 = \mathsf{M}\xi.$$

For $s > 1$ it is natural to consider, besides the moments $\alpha_s$ and $\beta_s$, the *central moments*

$$\mu_s = \mathsf{M}(\xi - \mathsf{M}\xi)^s = \int (x - \alpha_1)^s \, dF_\xi(x) \tag{7}$$

and the *central absolute moments*

$$\nu_s = \mathsf{M}|\xi - \mathsf{M}\xi|^s = \int |x - \alpha_1|^s \, dF_\xi(x). \tag{8}$$

From the inequalities

$$|\xi|^s \leqslant 2^s\{|\xi - \alpha_1|^s + |\alpha_1|^s\},$$
$$|\xi - \alpha_1|^s \leqslant 2^s\{|\xi|^s + |\alpha_1|^s\}$$

it is easy to deduce that the existence of the moments $\mu_s$ and $\nu_s$ is equivalent to the existence of $\alpha_s$ and $\beta_s$. Hence in the following we shall speak of the condition of existence of moment of order $s$, without specifying which one.

It is easily computed that the moments $\alpha_s$ and $\mu_s$ are connected by the relations

$$\left.\begin{aligned}
&\mu_0 = 1,\\
&\mu_1 = 0,\\
&\mu_2 = \alpha_2 - \alpha_1^2,\\
&\mu_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3,\\
&\mu_4 = \alpha_4 - 4\alpha_1\alpha_3 + 6\alpha_1^2\alpha_2 - 3\alpha_1^4,\\
&\dots\dots\dots\dots\dots\dots
\end{aligned}\right\} \tag{9}$$

These relations can be prolonged to any $s$ and yield the inversion

$$\left.\begin{aligned} \alpha_0 &= 1, \\ \alpha_1 &= \alpha_1, \\ \alpha_2 &= \mu_2 + \alpha_1^2, \\ \alpha_3 &= \mu_3 + 3\alpha_1\mu_2 + \alpha_1^3, \\ \alpha_4 &= \mu_4 + 4\alpha_1\mu_3 + 6\alpha_1^2\mu_2 + \alpha_1^4, \\ &\cdots\cdots\cdots\cdots\cdots \end{aligned}\right\} \tag{10}$$

The connection between the characteristic function and the moments is given by:

LEMMA 2. *If the random variable $\xi$ has a moment of order $k$, then its characteristic function $f(t)$ has continuous derivatives up to and including the $k$th order. Moreover*

$$\alpha_s = \frac{1}{i^s}\left[\frac{d^s}{dt^s} f_\xi(t)\right]_{t=0} \quad (s = 1, 2, \ldots, k). \tag{11}$$

*Proof.* By definition,

$$f_\xi(t) = \int e^{itx}\, dF_\xi(x).$$

Differentiate this equation formally $k$ times:

$$\frac{d^k}{dt^k} f_\xi(t) = i^k \int x^k e^{itx}\, dF_\xi(x). \tag{12}$$

Since by assumption

$$\int |x|^k\, dF_\xi(x) < \infty,$$

the integral on the right side of (12) exists and (12) can be proved by means of repeated integration with respect to $t$ (inverting the order of integrations with respect to $t$ and $x$ in accordance with Fubini's theorem). This proves the legitimacy of the differentiation. Putting $t = 0$ in (12), we arrive at (11).

Letting

$$\eta = \xi - \alpha_1,$$

we obtain for the central moments $\mu_s$ of the variable $\xi$ the expression

$$\mu_s = \mathsf{M}\eta^s = \frac{1}{i^s}\left[\frac{d^s}{dt^s} f_\eta(t)\right]_{t=0}.$$

Since

$$f_\eta(t) = e^{-it\alpha_1} f_\xi(t),$$

we have

$$\mu_s = \frac{1}{i^s}\left[\frac{d^s}{dt^s} e^{-it\alpha_1} f_\xi(t)\right]_{t=0}. \tag{13}$$

Comparison of (11) and (13) permits a new derivation of the relations (9) and (10).

The moments of the sum of two independent random variables

$$\xi = \xi^{(1)} + \xi^{(2)}$$

can be calculated:

$$\left.\begin{aligned}\alpha_1 &= \alpha_1^{(1)} + \alpha_1^{(2)},\\ \alpha_2 &= \alpha_2^{(1)} + 2\alpha_1^{(1)}\alpha_1^{(2)} + \alpha_2^{(2)},\\ \alpha_3 &= \alpha_3^{(1)} + 3\alpha_2^{(1)}\alpha_1^{(2)} + 3\alpha_1^{(1)}\alpha_2^{(2)} + \alpha_3^{(2)},\\ &\cdots\cdots\cdots\cdots\cdots\end{aligned}\right\} \tag{14}$$

$$\left.\begin{aligned}\mu_2 &= \mu_2^{(1)} + \mu_2^{(2)},\\ \mu_3 &= \mu_3^{(1)} + \mu_3^{(2)},\\ \mu_4 &= \mu_4^{(1)} + 6\mu_2^{(1)}\mu_2^{(2)} + \mu_4^{(2)},\\ &\cdots\cdots\cdots\cdots\end{aligned}\right\} \tag{15}$$

From these formulas we see that

$$\alpha_1 = \mathsf{M}\xi, \quad \mu_2 = \mathsf{D}^2\xi \text{ and } \mu_3$$

are additive for independent summands. In § 11 the question was raised whether there exist other numerical characteristics of distributions which possess this property. The answer is again essentially contained in the fundamental property of the characteristic function

$$f_\xi(t) = f_{\xi_1}(t) \cdot f_{\xi_2}(t),$$

from which it follows that

$$\log f_\xi(t) = \log f_{\xi_1}(t) + \log f_{\xi_2}(t). \tag{16}$$

Here, as everywhere in the sequel, $\log f(t)$ denotes the *principal branch* of the logarithm of the characteristic function, i.e., the function which is defined only for those real $t$ for which $f(t)$ is different from 0 both at the point $t$ and between $t$ and 0, and which is continuous and reduces at $t = 0$ to

$$\log 1 = 0.$$

As is easily verified, the principal branch of the logarithm of $f(t)$ is determined uniquely by these conditions. In this section we need only values of $\log f(t)$ in the neighborhood of zero.

If $\xi$ has a moment of the $s$th order, by Lemma 2 the first $s$ derivatives of both $f_\xi(t)$ and $\log f_\xi(t)$ exist at $t = 0$. Letting

$$\varkappa_r = \frac{1}{i^r}\left[\frac{d^r}{dt^r}\log f_\xi(t)\right]_{t=0},$$

we can therefore write

$$\log f_\xi(t) = \sum_{r=1}^{s} \frac{\varkappa_r}{r!}(it)^r + o(t^s). \tag{17}$$

From (16) and (17) it follows that for the sum of independent variables the corresponding coefficients $\chi_r$ add up:

$$\varkappa_r = \varkappa_r^{(1)} + \varkappa_r^{(2)}. \tag{18}$$

The coefficients $\chi_r$ in (17) are called the *semi-invariants* of the random variable $\xi$ (or of its distribution). The semi-invariants up to and including the order $s$ exist, if the moment $\beta_s$ exists.

The semi-invariants up to the $s$th order are uniquely determined by the moments up to the same order. Putting

$$w = it$$

and noting that by (11),

$$f_\xi(t) = \sum_{r=0}^{s} \frac{\alpha_r}{r!}(it)^r + o(t^s) \tag{19}$$

provided the moment $\alpha_s$ exists, we can write the relations between the semi-invariants and the moments in the form of an equation between formal power series:

$$\sum_{r=1}^{\infty} \frac{\chi_r}{r!} w^r = \log \sum_{r=0}^{\infty} \frac{\alpha_r}{r!} w^r. \tag{20}$$

This gives

$$\left.\begin{aligned} &\chi_1 = \alpha_1 = \mathsf{M}\xi, \\ &\chi_2 = \alpha_2 - \alpha_1^2 = \mathsf{D}^2\xi, \\ &\chi_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3, \\ &\chi_4 = \alpha_4 - 3\alpha_2^2 - 4\alpha_1\alpha_3 + 12\alpha_1^2\alpha_2 - 6\alpha_1^4, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}\right\} \tag{21}$$

and

$$\left.\begin{aligned} &\alpha_1 = \chi_1, \\ &\alpha_2 = \chi_2 + \chi_1^2, \\ &\alpha_3 = \chi_3 + 3\chi_1\chi_2 + \chi_1^3, \\ &\alpha_4 = \chi_4 + 3\chi_2^2 + 4\chi_1\chi_3 + 6\chi_1^2\chi_2 + \chi_1^4 \\ &\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}\right\} \tag{22}$$

From (17) and (19) we deduce that, for any $r \leq s$, $\chi_r/r!$ is the coefficient of $z^r$ in the expansion of $\log\left(1 + \sum_{k=1}^{r} \frac{\alpha_k}{k!} z^k\right)$ as a power series in $z$. By (5) this is majorized by the series

$$-\log\left[1 - \sum_{k=1}^{\infty} \frac{(\beta_r^{\frac{1}{r}} z)^k}{k!}\right] = \sum_{k=1}^{\infty} \frac{1}{k}\left(e^{\beta_r^{\frac{1}{r}} z} - 1\right)^k.$$

Consequently,

$$\frac{|\chi_r|}{r!} \leq \sum_{k=1}^{r} \frac{1}{k} \frac{k^r \beta_r}{r!} \leq \frac{r^r \beta_r}{r!},$$

that is to say,

$$|\varkappa_r| \leqslant r^r \beta_r. \tag{23}$$

For later purposes it is expedient to make use of the formula

$$e^{-it\alpha_1} f_\xi(t) = 1 + \sum_{r=2}^{s} \frac{\mu_r}{r!} (it)^r + o(t^s), \tag{24}$$

and to write down the connection between the semi-invariants and the central moments as a formal equation:

$$\sum_{r=1}^{\infty} \frac{\varkappa_r}{r!} w^r = \alpha_1 w + \log \left[ 1 + \sum_{r=2}^{\infty} \frac{\mu_r}{r!} w^r \right]. \tag{25}$$

This gives

$$\left.\begin{aligned}
&\varkappa_1 = \alpha_1 = \mathsf{M}\xi, \\
&\varkappa_2 = \mu_2 = \mathsf{D}^2\xi, \\
&\varkappa_3 = \mu_3, \\
&\varkappa_4 = \mu_4 - 3\mu_2^2, \\
&\varkappa_5 = \mu_5 - 10\mu_2\mu_3, \\
&\varkappa_6 = \mu_6 - 15\mu_2\mu_4 - 10\mu_3^2 + 30\mu_2^2, \\
&\cdots\cdots\cdots\cdots\cdots
\end{aligned}\right\} \quad (26$$

If we count the semi-invariants as the basic characteristics of probability distributions, then the simplest distribution with given

$$\varkappa_1 = \mathsf{M}\xi, \quad \varkappa_2 = \mathsf{D}^2\xi$$

should be the distribution with the characteristic function

$$f(t) = e^{i\varkappa_1 t - \frac{\varkappa_2}{2} t^2},$$

i.e., the normal distribution with the probability density

$$p(x) = \frac{1}{\sqrt{2\pi\varkappa_2}} e^{-\frac{(x-\varkappa_1)^2}{2\varkappa_2}}.$$

## CHAPTER 3

## INFINITELY DIVISIBLE DISTRIBUTIONS

### § 16. Statement of the Problem. Random Functions with Independent Increments

Classical limit theorems, the generalization and strengthening of which constitute the principal content of this book, have to do with sums

$$\zeta_n = \xi_1 + \xi_2 + \cdots + \xi_n$$

of an increasing number of terms of a sequence

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

of independent random variables.

It is possible to imagine an analogous scheme, in which the index $n$, taking only integral values, is replaced by a continuously varying parameter $\lambda$. There are no longer any elementary summands $\xi_n$, but it is possible to carry over the requirement, which in the discrete case follows from the independence of the variables $\xi_n$, that

$$\zeta_{n_2} - \zeta_{n_1}, \ \zeta_{n_3} - \zeta_{n_2}, \ldots, \zeta_{n_k} - \zeta_{n_{k-1}}$$

are independent if $n_1 < n_2 < \cdots < n_k$.

We confine ourselves to the continuous analogue of the case in which all elementary summands $\xi_n$ have the same law of distribution. Then the laws of distribution of the increments $\zeta_m - \zeta_n$ depend only on the number of summands entering into them, i.e., on the value of the difference $m - n$.

Our continuous scheme will look like this: to each real $\lambda \geq 0$ corresponds a random variable $\zeta_\lambda$ such that

(1) $\zeta_0$ is identically zero;

(2) the law of distribution of the difference $\zeta_{\lambda_2} - \zeta_{\lambda_1}$, with $\lambda_2 > \lambda_1$, depends only on the difference $\lambda_2 - \lambda_1$;

(3) for $\lambda_1 < \lambda_2 < \cdots \lambda_k$ the differences

$$\zeta_{\lambda_2} - \zeta_{\lambda_1}, \ \zeta_{\lambda_3} - \zeta_{\lambda_2}, \ \ldots, \ \zeta_{\lambda_k} - \zeta_{\lambda_{k-1}}$$

are mutually independent.

The problem of this chapter consists in the study of those distributions of the variable $\zeta_\lambda$ which are consistent with the scheme presented here. From the conditions (1), (2), and (3) it follows that $\zeta_\lambda$, for any natural number $\lambda = n$, is the sum

$$\zeta_\lambda = \eta_1 + \eta_2 + \cdots + \eta_n$$

of $n$ identically distributed independent summands

$$\eta_k = \zeta_{\frac{k}{n}\lambda} - \zeta_{\frac{k-1}{n}\lambda}.$$

This circumstance forms the basis of the formal definition of "infinitely divisible distribution" in § 17.

In Chapter 4 it will appear that infinitely divisible distributions play a fundamental role even in the classical problems of limit theorems for discrete sums of independent random variables. In the later sections of this chapter the properties of infinitely divisible distributions will be studied, by preference, purely analytically, by means of the characteristic function. However, almost all the results of this chapter were first discovered heuristically starting from the preceding scheme of the *random function* $\zeta_\lambda$, with independent increments, depending on the continuous argument $\lambda$.

Before the construction of the general theory two basic elementary types of such random functions were known:

(1) The normal type, in which the characteristic function $f_\lambda(t)$ of the random variable $\zeta_\lambda$ is given by the formula

$$\log f_\lambda(t) = \lambda\left(i\gamma t - \frac{\sigma^2}{2}t^2\right). \tag{1}$$

(2) The Poisson type, in which the characteristic function $f_\lambda(t)$ has the form

$$\log f_\lambda(t) = \lambda c\,(e^{iht} - 1). \tag{2}$$

It is possible to show that the normal type is the only one that can arise when $\zeta_\lambda$, as a function of $\lambda$, is continuous with probability one (see § 26) (a somewhat weaker proposition is proved in § 2 of Glivenko's book [32]). The Poisson type arises when $\zeta_\lambda$ as a function of $\lambda$ is, with probability one, a nondecreasing step function taking only values which are multiples of the "span" $h$ (this case of discrete jump-like random process is developed in § 2 of Glivenko's book [32]).

It is natural to try to build up a function $\zeta_\lambda$ combining these two types of variation and admitting not only jumps of a fixed magnitude $h$, but of all sorts of magnitudes. Let us, then, suppose that in the interval $(\lambda; \lambda + d\lambda)$ a jump occurs with probability $c\,d\lambda$, and that the distribution function of the magnitude of the jump is

$$\mathsf{P}(h < u) = F(u).$$

Then by combining (1) and (2) we arrive at the formula proposed first by de Finetti (see [30]):

$$\log f_\lambda(t) = \lambda\left\{i\gamma t - \frac{\sigma^2}{2}t^2 + c\int (e^{iut} - 1)\,dF(u)\right\}. \tag{3}$$

The formula (3), however, by no means yet gives the general solution of the question. In case $\zeta_\lambda$ has a finite variance the general solution was found by A. N. Kolmogorov (see later § 18). To this end two difficulties had to be overcome. First of all it is necessary to take into account the fact that jumps of small magnitudes can occur very often, and the full "density of jumps" may be infinite. Since jumps with large absolute values * cannot occur with infinite "density," it turned out to be possible to introduce two functions $M(u)$ and $N(u)$ such that in the interval $(\lambda, \lambda + d\lambda)$ the jumps $h$,

$$h < u < 0,$$

occur with probability $M(u)\, d\lambda$, and the jumps $h$,

$$h > u > 0,$$

with probability $N(u)\, d\lambda$. For $u = 0$ both these functions may become infinite. (3) now becomes

$$\log f_\lambda(t) = \lambda \left\{ i\gamma t - \frac{\sigma^2}{2} t^2 + \int_{-\infty}^{0} (e^{iut} - 1)\, dM(u) \right.$$

$$\left. + \int_{0}^{\infty} (e^{iut} - 1)\, dN(u) \right\}, \qquad (4)$$

where the integrals are understood to be the limits †

$$\int_{-\infty}^{0} (e^{iut} - 1)\, dM(u) = \lim_{a \to 0} \int_{-\infty}^{a} (e^{iut} - 1)\, dM(u), \quad a < 0,$$

$$\int_{0}^{\infty} (e^{iut} - 1)\, dN(u) = \lim_{a \to 0} \int_{a}^{\infty} (e^{iut} - 1)\, dN(u), \quad a > 0.$$

The second difficulty consists in that it is possible to have a function $\zeta_\lambda$ for which the integrals

$$\int_{-a}^{0} u\, dM(u) \quad \text{and} \quad \int_{0}^{a} u\, dN(u), \quad a > 0,$$

---

* *Translator's note.* That is, with absolute values bounded away from zero.

† In connection with the integral

$$\int_{a}^{\infty} (e^{iut} - 1)\, dN(u), \qquad a > 0,$$

it should be remarked that it is understood to be

$$\int_{u > a} (e^{iut} - 1)\, \mu(du),$$

where the measure $\mu(A)$ is determined by the condition

$$\mu(u;\, \infty) = -N(u).$$

and consequently also the integrals in (4), are divergent. This means that the mathematical expectation of the sum of the jumps of small magnitudes $|h| \leq a$ can be infinite. Roughly speaking, such an infinity can be compensated by introducing in the expression for $\log f_\lambda(t)$ a term $\lambda i\gamma t$ with an infinite value $\gamma$. This compensation is made rigorous by introducing a term proportional to $it$ inside those integrals whose divergence must be compensated. This leads to the formula

$$\log f_\lambda(t) = \lambda\left\{i\gamma t - \frac{\sigma^2}{2}t^2 + \int_{-\infty}^{0}(e^{iut} - 1 - itu)\,dM(u) + \int_{0}^{\infty}(e^{iut} - 1 - itu)\,dN(u)\right\}, \quad (5)$$

which already represents the general form of the logarithm of the characteristic function for variables $\zeta_\lambda$ of finite variance.

To cover the case of infinite variance, the correction term under the integral sign must be introduced with greater care. This was done by P. Lévy, who gave for $\log f_\lambda(t)$ the formula which is valid in the general case:

$$\log f_\lambda(t) = \lambda\left\{i\gamma t - \frac{\sigma^2}{2}t^2 + \int_{-\infty}^{0}\left(e^{iut} - 1 - \frac{iut}{1+u^2}\right)dM(u) + \int_{0}^{\infty}\left(e^{iut} - 1 - \frac{iut}{1+u^2}\right)dN(u)\right\}. \quad (6)$$

*The formula of Lévy and Khintchine,*

$$\log f_\lambda(t) = \lambda\left\{i\gamma t + \int\left(e^{itu} - 1 - \frac{itu}{1+u^2}\right)\frac{1+u^2}{u^2}\,dG(u)\right\}, \quad (7)$$

is obtained from (6) by introducing the function $G(u)$ with the properties

(1) $\dfrac{1+u^2}{u^2}dG(u) = dM(u)$ for $u < 0$;

(2) $\dfrac{1+u^2}{u^2}dG(u) = dN(u)$ for $u > 0$.

(3) The jump of $G(u)$ at $u = 0$ is equal to $\sigma^2$, and the integrand at zero is defined by continuity:

$$\left[\left(e^{itu} - 1 - \frac{itu}{1+u^2}\right)\frac{1+u^2}{u^2}\right]_{u=0} = -\frac{t^2}{2}.$$

The function $G(u)$ does not have a simple intuitive meaning, but it is more convenient to use in proofs than are $M(u)$ and $N(u)$.

## § 17. Definition and Basic Properties

We shall say that the random variable $\xi$ is *infinitely divisible* if for every natural number $n$ it can be represented as the sum

$$\xi = \xi_{n1} + \xi_{n2} + \dots + \xi_{nn}$$

of $n$ independent identically distributed random variables $\xi_{n1}, \xi_{n2}, \dots, \xi_{nn}$.

The distribution functions of infinitely divisible random variables will be called infinitely divisible distribution functions.

Obviously, the distribution function $F(x)$ is infinitely divisible if † and only if its characteristic function $f(t)$ is, for every natural number $n$, the $n$th power of some characteristic function $f_n(t)$ (which depends, of course, on $n$):

$$f(t) = [f_n(t)]^n. \tag{1}$$

The formula

$$f_n(t) = \sqrt[n]{f(t)}$$

does not yet uniquely determine the values of $f_n(t)$ in terms of the values of $f(t)$ (the $n$th root has $n$ values). But the additional requirements

(1) $f_n(0) = 1$,

(2) $f_n(t)$ is continuous

make it possible to determine $f_n(t)$ uniquely in every interval of $t$ containing the point $t = 0$ in which $f(t)$ does not vanish. We shall deal only with such a principal branch $\sqrt[n]{f(t)}$ in what follows. It will soon appear that the characteristic function $f(t)$ of an infinitely divisible law never vanishes (for real $t$). Therefore $f_n(t)$ and its corresponding distribution function $F_n(x)$ are uniquely determined by $f(t)$ (or $F(x)$).

We shall cite some examples.

Example 1. *A normally distributed random variable $\xi$ is infinitely divisible.*

In fact, suppose that $\mathbf{M}\xi = a$ and $\mathbf{D}\xi = \sigma^2$; then we know that the characteristic function of $\xi$ is

$$f(t) = e^{iat - \frac{\sigma^2}{2} t^2}.$$

Since for every $n > 0$

$$f_n(t) = e^{i\frac{a}{n}t - \frac{\sigma^2}{2n}t^2}$$

is the characteristic function of a normal law, our assertion is proved.

Example 2. *A random variable $\xi$ distributed according to a Poisson law is infinitely divisible.*

† *Translator's note.* See, however, pp. 247–248.

Suppose that the possible values of $\xi$ have the form $a + kh$ ($k = 0,1,2, \ldots$), and that

$$\mathsf{P}\{\xi = a + kh\} = \frac{\lambda^k e^{-\lambda}}{k!} \qquad (\lambda > 0).$$

Then the characteristic function of $\xi$ is

$$f(t) = e^{iat + \lambda(e^{ith} - 1)}.$$

From this we conclude that for every $n > 0$

$$\sqrt[n]{f(t)} = e^{i\frac{a}{n}t + \frac{\lambda}{n}(e^{ith} - 1)}$$

is the characteristic function of a random variable which is also distributed according to a Poisson law. This proves our assertion.

EXAMPLE 3. *A random variable distributed according to a Cauchy law* *

$$F(x) = \frac{1}{\pi}\left(\frac{\pi}{2} + \operatorname{arctg}\frac{x-b}{a}\right) \qquad (a > 0)$$

*is infinitely divisible.*

In fact, it can be calculated that the corresponding characteristic function is

$$f(t) = e^{ibt - a|t|}.$$

This proves our assertion.

EXAMPLE 4. *A random variable $\xi$ with the probability density*

$$p(x) = \begin{cases} 0 & \text{for } x \leqslant 0, \\ \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{for } x > 0, \end{cases}$$

*where $\alpha > 0$, $\beta > 0$ are constants, is infinitely divisible.*

The characteristic function of $\xi$ is

$$f(t) = \left(1 - \frac{it}{\beta}\right)^{-\alpha}.$$

For every natural number $n$

$$\sqrt[n]{f(t)} = \left(1 - \frac{it}{\beta}\right)^{-\frac{\alpha}{n}} \tag{2}$$

is again the characteristic function of a distribution of the same form as the initial one.

This distribution belongs to the system of Pearson curves of the third and the tenth types. It is known in statistics as the $\chi^2$ distribution, if $\beta = 1$ and $2\alpha$ is an integer.

THEOREM 1. *The characteristic function of an infinitely divisible law never vanishes.*

* The Cauchy law appeared for the first time in [14].

*Proof.* The assertion of the theorem is an obvious consequence of (1) and Theorem 4, § 14.

It is easy to verify that there exist any number of characteristic functions which do not vanish but at the same time are not infinitely divisible. For example, consider the discrete random variable taking the values $-1, 0, 1$ with the probabilities $\frac{1}{8}, \frac{3}{4}, \frac{1}{8}$. Its characteristic function

$$f(t) = \frac{1}{8} e^{-it} + \frac{3}{4} e^{it\,0} + \frac{1}{8} e^{it} = \frac{3 + \cos t}{4}$$

is positive and therefore does not vanish. Not only is the variable not infinitely divisible, but it cannot be represented as the sum of two identically distributed independent variables. In fact, suppose that

$$\xi = \xi_1 + \xi_2,$$

where $\xi_1$ and $\xi_2$ are mutually independent and identically distributed. It is evident that each of the summands can take only two values $a_1$ and $a_2$ $(a_1 < a_2)$, with probabilities $p$ and $q = 1 - p$ respectively. The possible values of $\xi_1 + \xi_2$ are the numbers $2a_1$, $a_1 + a_2$, and $2a_2$. The probabilities of these values are, of course, equal to $p^2$, $2pq$, and $q^2$. Since, by hypothesis, $2a_1 = -1$, $a_1 + a_2 = 0$, $2a_2 = 1$, while $\frac{1}{8} = p^2$, $\frac{3}{4} = 2pq$, and $\frac{1}{8} = q^2$, we reach a contradiction, for the last three equations are inconsistent.

THEOREM 2. *The distribution function of the sum of a finite number of independent infinitely divisible random variables is itself infinitely divisible.*

*Proof.* Obviously, it is sufficient in the proof to confine ourselves to the case of two summands $\xi$ and $\eta$. If $f(t)$ and $g(t)$ are the characteristic functions of these variables, then, by the condition of the theorem for every natural number $n$, we have

$$f(t) = \{f_n(t)\}^n, \quad g(t) = \{g_n(t)\}^n,$$

where $f_n(t)$ and $g_n(t)$ are characteristic functions. The characteristic function $h(t)$ of the sum $\zeta = \xi + \eta$ satisfies the equation

$$h(t) = f(t) \cdot g(t) = \{f_n(t) \cdot g_n(t)\}^n$$

for every $n$, which obviously proves the theorem.

We remark that the converse proposition is not true and that it is possible to give examples of random variables which are not infinitely divisible but whose sum is infinitely divisible. We postpone the consideration of such examples until the next section (Example 1).

THEOREM 3. *A distribution function which is the limit, in the sense of weak convergence, of infinitely divisible distribution functions is itself infinitely divisible.*

*Proof.* Let $F^{(k)}(x)$ be an infinitely divisible distribution function and let

$$F^{(k)}(x) \Rightarrow F(x)$$

as $k \to \infty$, where $F(x)$ is a distribution function. If $f^{(k)}(t)$ is the characteristic function of $F^{(k)}(x)$, and $f(t)$ is that of $F(x)$, then

$$f^{(k)}(t) \Rightarrow f(t). \tag{3}$$

By the condition of the theorem, for every $n$ the function

$$f_n^{(k)}(t) = \sqrt[n]{f^{(k)}(t)}$$

is a characteristic function and never vanishes for any $t$. From (3) we therefore conclude that for every $n$

$$f_n^{(k)}(t) \Rightarrow f_n(t), \qquad k \to \infty.$$

From Theorem 2 of § 13 it follows that $f_n(t)$ is a characteristic function. Since for every natural number $n$ the equation

$$f(t) = \{f_n(t)\}^n$$

holds, the theorem is proved.

THEOREM 4. *If $f(t)$ is the characteristic function of an infinitely divisible distribution function, then for every $c > 0$ the function $\{f(t)\}^c$ is also a characteristic function.*

*Proof.* In fact, for $c = 1/n$, when $n$ is a natural number, this follows from the definition of an infinitely divisible random variable. By the theorem on the multiplication of characteristic functions this assertion remains true for any rational number $c > 0$. Finally, for an irrational number $c > 0$ the function $\{f(t)\}^c$ can be approximated uniformly in every finite interval $t$ by the function $\{f(t)\}^{c_1}$, where $c_1$ is a rational number. Hence our assertion follows from a preceding theorem.

THEOREM 5. *The totality of infinitely divisible distribution laws coincides with the totality of laws which are composed of a finite number of Poisson laws and of limits of these laws in the sense of weak convergence.*

*Proof.* That the composition of a finite number of Poisson laws and their limit laws is infinitely divisible, follows from Theorems 2 and 3. We shall prove the converse. Let $f(t)$ be the characteristic function of an infinitely divisible law. By hypothesis,

$$f_n(t) = \sqrt[n]{f(t)}$$

is a characteristic function; thus

$$f_n(t) = \int e^{itx}\, dF_n(x), \tag{4}$$

where $F_n(x)$ is a distribution function. We have *

* This relation can be proved, e.g., in the following way:

$$n\left(\sqrt[n]{a} - 1\right) = n\left(e^{\frac{1}{n}\log a} - 1\right) = n\left(1 + \frac{1}{n}\log a + o\left(\frac{1}{n}\right) - 1\right) \to \log a.$$

$$n(f_n(t) - 1) \Rightarrow \log f(t) \qquad \text{as } n \to \infty \tag{5}$$

and therefore

$$e^{n(f_n(t)-1)} \Rightarrow f(t), \qquad n \to \infty. \tag{6}$$

We represent the integral (4) as the limit as $m \to \infty$ of the Stieltjes sum:

$$\sum_{k=1}^{m} e^{itc_k}[F_n(c_k) - F_n(c_{k-1})] \Rightarrow f_n(t). \tag{7}$$

Put

$$a_k = n[F_n(c_k) - F_n(c_{k-1})].$$

Comparing (6) and (7), it is easy to see that

$$e^{\sum_{k=1}^{m} a_k(e^{itc_k}-1)} \Rightarrow f(t). \tag{8}$$

Q.E.D.

We shall make use of the last theorem to construct examples of infinitely divisible distributions.

EXAMPLE 5. *The function*

$$f(t) = (1 - b) \cdot (1 - be^{it})^{-1} \quad (0 < b < 1)$$

*is the characteristic function of an infinitely divisible distribution.*

First of all, from the equation

$$f(t) = (1 - b) \sum_{n=0}^{\infty} b^n e^{int}$$

we conclude that $f(t)$ is the characteristic function of a random variable $\xi$ which takes only non-negative integral values with the probabilities

$$\mathsf{P}\{\xi = n\} = (1 - b)b^n \qquad (n = 0, 1, 2, \ldots).$$

It is easily calculated that

$$\log f(t) = \sum_{k=1}^{\infty} (e^{ikt} - 1)\frac{b^k}{k}.$$

Since each separate term of this sum is the logarithm of the characteristic function of a Poisson law, the assertion is proved.

EXAMPLE 6.* *Let $\zeta(s) = \zeta(\sigma + it)$ be the Riemann zeta-function defined for $\sigma > 1$ by means of the series*

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$$

*or the Euler product*

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1},$$

*extended over all prime numbers.*

---

* A. Ya. Khintchine [59], p. 35.

We shall prove that for every $\sigma > 1$ the function

$$f(t) = \frac{\zeta(\sigma + it)}{\zeta(\sigma)}$$

is the characteristic function of an infinitely divisible distribution. In fact

$$\log f(t) = \sum_p [\log(1 - p^{-\sigma}) - \log(1 - p^{-\sigma - it})]$$

$$= \sum_p \sum_{m=1}^{\infty} \frac{p^{-m\sigma}(p^{-imt} - 1)}{m} = \sum_p \sum_{m=1}^{\infty} \frac{p^{-m\sigma}(e^{-imt \log p} - 1)}{m},$$

where the symbol $\Sigma_p$ denotes that the summation is extended over all the prime numbers.

Each term of this sum is the logarithm of the characteristic function of a Poisson law. According to Theorem 5 the characteristic function $f(t)$ is infinitely divisible.

## § 18. The Canonical Representation

THEOREM 1. *In order that the function $f(t)$ be the characteristic function of an infinitely divisible distribution, it is necessary and sufficient that its logarithm be representable in the form*

$$\log f(t) = i\gamma t + \int \left\{ e^{itu} - 1 - \frac{itu}{1 + u^2} \right\} \frac{1 + u^2}{u^2} \, dG(u), \qquad (1)$$

*where $\gamma$ is a real constant, $G(u)$ is a nondecreasing function of bounded variation, and the integrand at $u = 0$ is defined by the equation*

$$\left[ \left\{ e^{itu} - 1 - \frac{itu}{1 + u^2} \right\} \frac{1 + u^2}{u^2} \right]_{u=0} = -\frac{t^2}{2}.$$

*The representation of $\log f(t)$ by the formula* (1) *is unique.*

*Proof. Necessity.* Suppose that $F(x)$ is an infinitely divisible distribution and $f(t)$ its characteristic function. Then for every $n > 0$

$$f(t) = [f_n(t)]^n,$$

where $f_n(t)$ is a characteristic function. Since $f(t) \neq 0$, according to (5) of § 17 we have

$$n[f_n(t) - 1] = n \int (e^{itx} - 1) \, dF_n(x) \Rightarrow \log f(t),$$

where $F_n(x)$ is the distribution function corresponding to the characteristic function $f_n(t)$. Put

$$G_n(u) = n \int_{-\infty}^{u} \frac{x^2}{1 + x^2} \, dF_n(x)$$

and

$$I_n(t) = \int (e^{itu} - 1) \frac{1+u^2}{u^2} dG_n(u). \tag{2}$$

Then by Theorem 2 of § 4 the preceding relation may be written as

$$I_n(t) \Rightarrow \log f(t) \tag{3}$$

and we conclude that *

$$\operatorname{Re} I_n(t) = \int (\cos ut - 1) \frac{1+u^2}{u^2} dG_n(u)$$
$$\Rightarrow \int (\cos ut - 1) \frac{1+u^2}{u^2} dG(u) = \log |f(t)|.$$

We shall prove that $G_n(+\infty)$ is bounded. For this purpose consider the expressions

$$A_n = \int\limits_{|u|\leq 1} dG_n(u),\ B_n = \int\limits_{|u|>1} dG_n(u),\ C_n = A_n + B_n = \int dG_n(u).$$

Let $0 \leq t \leq 2$. It is evident that for every $\epsilon > 0$ and for sufficiently large $n$

$$-\log|f(t)| + \varepsilon \geqslant \int\limits_{|u|\leqslant 1} (1 - \cos tu) \frac{1+u^2}{u^2} dG_n(u)$$

and

$$-\log|f(t)| + \varepsilon \geqslant \int\limits_{|u|>1} (1 - \cos tu) \frac{1+u^2}{u^2} dG_n(u).$$

For $|u| \leq 1$

$$\frac{1-\cos u}{u^2} > \frac{1}{3},$$

hence the first inequality above gives

$$-\log|f(1)| + \varepsilon > \frac{1}{3} A_n. \tag{4}$$

Taking, in the interval $0 \leq t \leq 2$, the mean of the functions on both sides of the second inequality above, we obtain

$$-\frac{1}{2}\int_0^2 \log|f(t)|\,dt + \varepsilon \geqslant \int\limits_{|u|>1} \left(1 - \frac{\sin 2u}{2u}\right) dG_n(u) \geqslant \frac{1}{2} B_n. \tag{5}$$

Since the quantities $\log|f(1)|$ and $\frac{1}{2}\int_0^2 \log|f(t)|\,dt$ are finite, it follows

* *Translator's note.* The existence of a $G(u)$ such that $G_n(u) \Rightarrow G(u)$ will appear in the course of the proof. This fact, however, is not needed at all. Thus the second integral should be deleted from the formula below.

from (4) and (5) that $G_n(+\infty)$ is bounded. We shall now prove that

$$\int\limits_{|u|>T} dG_n(u) \to 0 \qquad \text{as } T \to \infty$$

uniformly with respect to $n$. In fact, for every $\epsilon > 0$ and for sufficiently large $n$

$$-\log|f(t)| + \varepsilon \geqslant \int\limits_{|u|\geqslant T} (1 - \cos tu)\, dG_n(u).$$

Taking the mean in the interval $0 \leq t \leq \frac{2}{T}$ $(T \geq 1)$ of both sides of the inequality, we obtain

$$-\frac{T}{2}\int_0^{\frac{2}{T}} \log|f(t)|\, dt + \varepsilon \geqslant \int\limits_{|u|\geqslant T} \left(1 - \frac{T \sin\frac{2u}{T}}{2u}\right) dG_n(u).$$

But for $|u| \geq T$

$$1 - \frac{T\sin\frac{2u}{T}}{2u} \geqslant \frac{1}{2},$$

and for $T \geq T_0$

$$\left|\frac{T}{2}\int_0^{\frac{2}{T}} \log|f(t)|\, dt\right| \leqslant \max_{0 \leqslant t \leqslant \frac{2}{T}} |\log|f(t)|| < \varepsilon.$$

Therefore for $T \geq T_0$

$$\int\limits_{|u|\geqslant T} dG_n(u) \leqslant 4\varepsilon.$$

Now on the basis of Theorem 3 bis of § 9 we can choose a subsequence from $G_n(u)$ such that

$$G_{n_k}(u) \Longrightarrow G(u),$$

where $G(u)$ is a nondecreasing function of bounded variation.*

Put

$$\gamma_{n_k} = \int \frac{dG_{n_k}(u)}{u} = n_k \int \frac{x}{1+x^2}\, dF_{n_k}(x);$$

then it is evident from (3) that

$$I_{n_k}(t) = \int \left\{ e^{itu} - 1 - \frac{itu}{1+u^2} \right\} \frac{1+u^2}{u^2}\, dG_{n_k}(u) + it\gamma_{n_k}.$$

---

* *Translator's note.* The last clause is added for the sake of clarity, cf. the preceding note.

The integral on the right side of this equation, as $k \to \infty$, converges to

$$\int \left\{ e^{itu} - 1 - \frac{itu}{1+u^2} \right\} \frac{1+u^2}{u^2}\, dG(u).$$

From (3) we conclude that $\gamma_{n_k}$ must converge to some number $\gamma$ as $k \to \infty$.

Thus the first part of the theorem is proved.

*Sufficiency.* Suppose that (1) holds. According to § 6 the integral on the right side of (1) is the limit of the Stieltjes sums (all $c_k$ are taken to be different from zero):

$$\sum_{k=1}^{m} \left( e^{itc_k} - 1 - \frac{itc_k}{1+c_k^2} \right) \frac{1+c_k^2}{c_k^2} [G(c_k) - G(c_{k-1})]$$

$$\Longrightarrow \int \left( e^{itu} - 1 - \frac{itu}{1+u^2} \right) \frac{1+u^2}{u^2}\, dG(u).$$

Each term on the left is the logarithm of the characteristic function of a Poisson law.

Applying Theorem 5 of § 17, we confirm that $f(t)$ is an infinitely divisible characteristic function.

It remains to prove the uniqueness of the representation of the logarithm of an infinitely divisible characteristic function by (1). From (1) it is easily deduced that

$$-v(t) = \int_{t-1}^{t+1} \log f(z)\, dz - 2 \log f(t)$$

$$= -2 \int e^{itu} \left( 1 - \frac{\sin u}{u} \right) \frac{1+u^2}{u^2}\, dG(u).$$

Putting

$$V(u) = 2 \int_{-\infty}^{u} \left( 1 - \frac{\sin v}{v} \right) \frac{1+v^2}{v^2}\, dG(v),$$

we find that

$$v(t) = \int e^{itu}\, dV(u).$$

The function $V(u)$ is nondecreasing, hence by Theorem 2 of § 12 it is uniquely determined by its characteristic function $v(t)$.

Since for all $v$

$$\left( 1 - \frac{\sin v}{v} \right) \frac{1+v^2}{v^2} > 0,$$

according to Theorem 3 of § 4 the function $G(v)$ is uniquely determined by the function $V(u)$. Q.E.D.

We shall call (1) the formula of Lévy and Khintchine as already indicated in § 16 (see [74] and [56]).

COROLLARY. *If the logarithm of a characteristic function is representable in the form (1), where $G(u)$ is a function of bounded variation (not necessarily nondecreasing), then such a representation is unique.*

*Proof.* Suppose that $f(t)$ has two representations of the form (1) with functions $G_1(u)$, $G_2(u)$ and constants $\gamma_1$, $\gamma_2$ respectively.

Let $G_1'(G_2')$ be the positive, $G_2''(G_2'')$ the negative variation of the function $G_1(G_2)$ and let $h(t,u)$ be the integrand in (1).

We obtain then [cf. (2), § 8]:

$$\begin{aligned} & i\gamma_1 t + \int h(t, u)\, dG_1(u) \\ & = i\gamma_1 t + \int h(t, u)\, dG_1'(u) - \int h(t, u)\, dG_1''(u) \\ & = i\gamma_2 t + \int h(t, u)\, dG_2(u) \\ & = i\gamma_2 t + \int h(t, u)\, dG_2'(u) - \int h(t, u)\, dG_2''(u), \end{aligned}$$

whence

$$\begin{aligned} & i\gamma_1 t + \int h(t, u)\, d\,(G_1'(u) + G_2''(u)) \\ & \qquad = i\gamma_2 t + \int h(t, u)\, d\,(G_1''(u) + G_2'(u)). \end{aligned}$$

But the functions $G_1'(u) + G_2''(u)$ and $G_1''(u) + G_2'(u)$ are nondecreasing; hence in the equation written above, *representing the logarithm of the characteristic function of an infinitely divisible law*, we must have

$$\gamma_1 = \gamma_2 \quad \text{and} \quad G_1'(u) + G_2''(u) \equiv G_1''(u) + G_2'(u),$$

that is,

$$G_1'(u) - G_1''(u) \equiv G_2'(u) - G_2''(u).$$

This is equivalent to

$$G_1 \equiv G_2,$$

Q.E.D.

We shall make use of the formula of Lévy and Khintchine to construct examples of infinitely divisible characteristic functions.

EXAMPLE 1. Consider the function

$$f(t)=\frac{1-\beta}{1+\alpha}\cdot\frac{1+\alpha e^{-it}}{1-\beta e^{it}}$$

$$(0<\alpha\leqslant\beta<1).$$

This function is continuous, $f(0) = 1$ and

$$f(t)=\frac{1-\beta}{1+\alpha}\left[\alpha e^{-it}+(1+\alpha\beta)\sum_{n=0}^{\infty}\beta^n e^{int}\right].$$

Hence it is the characteristic function of a random variable taking all integral values from $-1$ to $+\infty$ ; moreover

$$\mathsf{P}\{\xi=-1\}=\frac{1-\beta}{1+\alpha}\alpha;\ \ \mathsf{P}\{\xi=n\}=\frac{1-\beta}{1+\alpha}\cdot(1+\alpha\beta)\beta^n\ \ [n=0, 1,2, \ldots].$$

We observe that $f(t)$ is not an infinitely divisible characteristic function. For,

$$\log f(t)=\sum_{n=1}^{\infty}\left[(-1)^{n-1}\frac{\alpha^n}{n}(e^{-int}-1)+\frac{\beta^n}{n}(e^{int}-1)\right]$$

$$=i\gamma t+\int\left(e^{itu}-1-\frac{itu}{1+u^2}\right)\frac{1+u^2}{u^2}\,dG(u),$$

where, as can be easily calculated,

$$\gamma=\sum_{n=1}^{\infty}\frac{\beta^n+(-1)^n\alpha^n}{1+n^2},$$

and $G(u)$ is a function of bounded variation, having jumps at the points

$$u=\pm 1,\ \ \pm 2,\ \ \pm 3,\ \ldots,$$

of magnitudes

$$\frac{n\beta^n}{n^2+1}\quad\text{for}\quad u=+n$$

and

$$(-1)^{n-1}\frac{n\alpha^n}{n^2+1}\quad\text{for}\quad u=-n.$$

Thus $G(u)$ is *not monotone*. Hence according to the Corollary of Theorem 1 of this section, $f(t)$ cannot be the characteristic function of an infinitely divisible law, Q.E.D.

The function

$$\overline{f(t)}=\frac{1-\beta}{1+\alpha}\cdot\frac{1+\alpha e^{it}}{1-\beta e^{-it}}$$

is also a characteristic function, and

$$\log \overline{f(t)} = \sum_{n=1}^{\infty} \left[ \frac{\beta^n}{n} (e^{-int} - 1) + (-1)^{n-1} \frac{\alpha^n}{n} (e^{int} - 1) \right].$$

We shall prove that

$$g(t) = f(t)\overline{f(t)} = |f(t)|^2$$

is the characteristic function of an infinitely divisible law.

In fact,

$$\log g(t) = \sum_{n=1}^{\infty} \frac{1}{n} (\beta^n + (-1)^{n-1}\alpha^n)(e^{-int} - 1)$$

$$+ \sum_{n=1}^{\infty} \frac{1}{n} (\beta^n + (-1)^{n-1}\alpha^n)(e^{int} - 1)$$

$$= \int \left\{ e^{itx} - 1 - \frac{itx}{1+x^2} \right\} \frac{1+x^2}{x^2}\, dG(x),$$

where $G(x)$ is a nondecreasing function with jumps at the points $\pm 1$, $\pm 2$, $\pm 3, \ldots$; the jumps at the points $+n$ and $-n$ are equal; their magnitude is

$$\frac{n}{1+n^2} (\beta^n + (-1)^{n-1}\alpha^n)$$

for $n > 0$.

It is interesting to note the following: $f(t)$ is a characteristic function but is not infinitely divisible; its modulus $|f(t)|$ is an infinitely divisible characteristic function; the infinitely divisible characteristic function $|f(t)|^2$ is decomposed into the product of two characteristic functions $f(t)$ and $\overline{f(t)}$, neither of which is infinitely divisible.

We remark further that by our example we have also proved the assertion that there exist essentially different† characteristic functions whose moduli are the same.

EXAMPLE 2. It is easy to construct examples of infinitely divisible characteristic functions $f(t)$ with even more striking properties: the function $f(t)$ is decomposed into the product of an infinitely divisible characteristic function and two characteristic functions, neither of which is decomposable (A. Ya. Khintchine [57]).

The even function

$$\psi(t) = \log \frac{5 + 4\cos t}{9}$$

† *Translator's note.* Namely $f(t)$ and $|f(t)|$ in the example. Of course, for every characteristic function $f(t)$ there are trivially different characteristic functions with the same modulus, namely, $e^{i\alpha t} f(t)$ and $e^{i\alpha t} \overline{f(t)}$ for every real $\alpha$.

of the real argument $t$ has the period $2\pi$ and derivatives of all orders, hence

$$\psi(t)=\sum_{n=0}^{\infty} a_n \cos nt, \quad \sum_{n=0}^{\infty} |a_n| < +\infty.$$

Since $\psi(0)=0$,

$$a_0=-\sum_{n=1}^{\infty} a_n$$

and consequently

$$\psi(t)=\sum_{n=1}^{\infty} a_n(\cos nt-1).$$

Let $p_1, p_2, \ldots, p_k, \ldots$ be the non-negative and $-n_1, -n_2, \ldots, -n_k, \ldots$ the negative numbers among $a_1, a_2, \ldots, a_n, \ldots$. Thus

$$\psi(t)=\sum_{k=1}^{\infty} p_k(\cos l_k t-1)-\sum_{k=1}^{\infty} n_k(\cos m_k t-1),$$

whence

$$\prod_{k=1}^{\infty} e^{p_k(\cos l_k t-1)}=\frac{5+4\cos t}{9}\prod_{k=1}^{\infty} e^{n_k(\cos m_k t-1)}.$$

The infinite products on both sides of the equation, being the products of the moduli of the characteristic functions of Poisson laws, are themselves infinitely divisible characteristic functions.

In fact if $f(t)$ is an infinitely divisible characteristic function so is $|f(t)|^2$ and hence also $|f(t)|$. It remains to apply Theorems 2 and 3 of § 17. The function $\frac{5+4\cos t}{9}=\frac{2+e^{it}}{3}\cdot\frac{2+e^{-it}}{3}$ is the product of two characteristic functions, each of which is indecomposable. Thus it is possible to find an example of an infinitely divisible random variable represented as the sum of three independent random variables, one of which is infinitely divisible while the other two (nonconstant) are indecomposable into independent summands.

Define the functions $M(u)$ and $N(u)$ and the constant $\sigma^2$ by setting

$$\left.\begin{aligned} M(u)&=\int_{-\infty}^{u}\frac{1+z^2}{z^2}\,dG(z) \quad \text{for} \quad u<0,\\ N(u)&=-\int_{u}^{\infty}\frac{1+z^2}{z^2}\,dG(z) \quad \text{for} \quad u>0,\\ \sigma^2&=G(+0)-G(-0). \end{aligned}\right\} \tag{6}$$

The functions $M(u)$ and $N(u)$

(1) are respectively nondecreasing in the intervals

$$(-\infty, 0), \ (0, +\infty);$$

(2) are continuous at those and only those points at which $G(u)$ is continuous;

(3) satisfy the relations

$$M(-\infty) = N(+\infty) = 0$$

and

$$\int_{-\varepsilon}^{0} u^2 \, dM(u) + \int_{0}^{\varepsilon} u^2 dN(u) < +\infty$$

for every finite $\epsilon > 0$.

Conversely, any two functions $M(u)$ and $N(u)$ satisfying the conditions (1) and (3) and any constant $\sigma > 0$ determine by (6) the characteristic function of some infinitely divisible law. In terms of $M(u)$ and $N(u)$ we can write (1) in the following form [cf. (6), § 16]:

$$\log f(t) = i\gamma t - \frac{\sigma^2}{2} t^2 + \int_{-\infty}^{0} \left( e^{iut} - 1 - \frac{iut}{1+u^2} \right) dM(u) + \int_{0}^{\infty} \left( e^{iut} - 1 - \frac{iut}{1+u^2} \right) dN(u). \quad (7)$$

We shall call (7) *Lévy's formula.*

Finally, we can give (1) the following form:

$$\log f(t) = i\gamma(\tau) t - \frac{\sigma^2}{2} t^2 + \int_{-\infty}^{-\tau} (e^{iut} - 1) \, dM(u) + \int_{\tau}^{\infty} (e^{iut} - 1) \, dN(u) + \int_{-\tau}^{0} (e^{iut} - 1 - iut) \, dM(u) + \int_{0}^{\tau} (e^{iut} - 1 - iut) \, dN(u), \quad (8)$$

where $M(u)$ and $N(u)$ have the same meanings as in (6), and $\tau$ is an arbitrary constant, chosen so that $\tau$ and $-\tau$ are continuity points of the functions $N(u)$ and $M(u)$ respectively. The relation between $\gamma(\tau)$ and the $\gamma$ in the formula (1) is easily found:

$$\gamma(\tau) = \gamma + \int_{|u| < \tau} u \, dG(u) - \int_{|u| \geqslant \tau} \frac{1}{u} \, dG(u). \quad (9)$$

Lévy's formula and that of Lévy and Khintchine are generalizations of Kolmogorov's formula, which was found by him as early as 1932 [64], for infinitely divisible laws $F(x)$ of finite variance. It turns out that in that case $F(x)$ is infinitely divisible if and only if

$$\log f(t) = i\gamma t + \int \{e^{itu} - 1 - itu\} \frac{1}{u^2} dK(u), \tag{10}$$

where $\gamma$ is a constant and $K(u)$ $[K(-\infty) = 0]$ is a nondecreasing function of bounded variation. The representation of $\log f(t)$ by this formula is unique.

We shall call (10) *Kolmogorov's formula.* An easy calculation shows that in this case

$$\left[\frac{d}{dt}\log f(t)\right]_{t=0} = i\mathsf{M}\xi = i\gamma,$$

$$\left[\frac{d^2}{dt^2}\log f(t)\right]_{t=0} = -\mathsf{D}^2\xi = -\int dK(u),$$

whence

$$\gamma = \mathsf{M}\xi \quad \text{and} \quad K(+\infty) = \mathsf{D}^2\xi. \tag{11}$$

These formulas clarify the probability meanings of the constant $\gamma$ and the variation of $K(u)$ in (10).

Kolmogorov's formula can be obtained either from that of Lévy and Khintchine considering the function

$$K(u) = \int_{-\infty}^{u} (1 + v^2)\, dG(v),$$

or in the same way as the formula of Lévy and Khintchine itself was obtained.

For later purposes we shall need to know the representation of a normal law and a Poisson law by means of (1), (7), and (10). It is easily calculated that for the normal law

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(z-a)^2}{2\sigma^2}} dz \tag{12}$$

we should put in (1) and (10)

$$\gamma = a; \quad K(u) \equiv G(u) = \begin{cases} 0 & \text{for} \quad u \leqslant 0, \\ \sigma^2 & \text{for} \quad u > 0, \end{cases} \tag{13}$$

and in (7)

$$\gamma = a, \quad M(u) \equiv 0, \quad N(u) \equiv 0, \quad \sigma = \sigma. \tag{13'}$$

For a random variable distributed according to the Poisson law with the characteristic function

$$f(t) = e^{\lambda(e^{it} - 1)},$$

we should put in (1)

$$\gamma=\frac{\lambda}{2}, \quad G(u)=\begin{cases} 0 & \text{for} \quad u\leqslant 1, \\ \frac{\lambda}{2} & \text{for} \quad u>1; \end{cases} \tag{14}$$

in (7)

$$\gamma=\frac{\lambda}{2}, \quad \sigma=0, \quad M(u)\equiv 0, \quad N(u)=\begin{cases} -\lambda & \text{for} \quad u\leqslant 1, \\ 0 & \text{for} \quad u>1; \end{cases} \tag{14'}$$

and in (9)

$$\gamma=\lambda, \quad K(u)=\begin{cases} 0 & \text{for} \quad u\leqslant 1, \\ \lambda & \text{for} \quad u>1. \end{cases} \tag{14''}$$

We remark that the form of the function $G(u)$ in the formula of Lévy and Khintchine for a normal law and a Poisson law yields the following assertion:

If the composition of two infinitely divisible laws is a normal (Poisson) law, then each of the components must also be a normal (Poisson) law.

In fact, if $G_1(u)$ and $G_2(u)$ are two nondecreasing functions and

$$G_1(u)+G_2(u)=G(u),$$

where $G(u)$ is defined by (13) [or (14) in the case of a Poisson law], then both $G_1(u)$ and $G_2(u)$ can have only one point of increase at $u = 0$ (or at $u = 1$).

We shall now calculate the function $K(u)$ in Kolmogorov's formula for the distribution considered in Example 4, § 17. To this end we shall go through, for this example, all the computations made in the proof of (1). Thus we put

$$K_n(x)=n\int_{-\infty}^{x} z^2 p_n(z)\,dz,$$

where $p_n(x)$ is the density of the distribution whose characteristic function is determined by (1) of § 17, i.e.,

$$p_n(x)=\left\{\begin{matrix} 0 & \text{for} \quad x\leqslant 0, \\ c_n x^{\frac{\alpha}{n}-1}\, e^{-\beta x} & \text{for} \quad x>0, \end{matrix}\right\}$$

where $c_n=\dfrac{\beta^{\frac{\alpha}{n}}}{\Gamma\left(\frac{\alpha}{n}\right)}$.

Thus,

$$K_n(x)=\begin{cases} 0 & \text{for } x\leqslant 0, \\ nc_n\displaystyle\int_0^x z^{1+\frac{\alpha}{n}} e^{-\beta z}\,dz & \text{for } x>0. \end{cases}$$

Now we remark that

$$nc_n = \frac{\alpha\beta^{\frac{\alpha}{n}}}{\frac{\alpha}{n}\Gamma\left(\frac{\alpha}{n}\right)} = \frac{\alpha\beta^{\frac{\alpha}{n}}}{\Gamma\left(1+\frac{\alpha}{n}\right)} \to \alpha \quad \text{as} \quad n \to \infty$$

and for every $x > 0$

$$K_n(x) \to \alpha \int_0^x z e^{-\beta z}\, dz = K(x) \quad (n \to \infty).$$

Moreover, as $n \to \infty$

$$K_n(+\infty) = nc_n \int_0^\infty z^{1+\frac{\alpha}{n}} e^{-\beta z}\, dz = nc_n \frac{\Gamma\left(2+\frac{\alpha}{n}\right)}{\beta^{2+\frac{\alpha}{n}}} \to \frac{\alpha}{\beta^2} = K(+\infty).$$

We find also that

$$\gamma = \mathsf{M}\,\xi = c \int_0^\infty x^\alpha e^{-\beta x}\, dx = \frac{c\,\Gamma(\alpha+1)}{\beta^{\alpha+1}} = \frac{\alpha}{\beta}.$$

Thus, according to (10) we have the equation

$$\log f(t) = -\alpha \log\left(1 - \frac{it}{\beta}\right) = i\frac{\alpha}{\beta}t + \alpha \int_0^\infty \{e^{itx} - 1 - itx\} \frac{e^{-\beta x}}{x}\, dx.$$

## § 19. Conditions for Convergence of Infinitely Divisible Distributions

Theorem 1.* *For the convergence of a sequence $\{F_n(x)\}$ of infinitely divisible laws to a limit law $F(x)$ it is necessary and sufficient that as $n \to \infty$*

$$(1)\ G_n(x) \Rightarrow G(x),$$
$$(2)\ \gamma_n \to \gamma,$$

*where the functions $G_n(x)$ and $G(x)$ and the constants $\gamma_n$ and $\gamma$ are defined by the formula of Lévy and Khintchine for $F_n(x)$ and $F(x)$ respectively.*

*Proof. Necessity.* Suppose that $F_n(x) \Rightarrow F(x)$. Then $f_n(t) \Rightarrow f(t)$ according to Theorem 1 of § 13. Since $f_n(t)$ and $f(t)$ do not vanish for any $t$, as $n \to \infty$

$$l_n(t) = i\gamma_n t + \int \left(e^{iut} - 1 - \frac{iut}{1+u^2}\right)\frac{1+u^2}{u^2}\, dG_n(u)$$
$$\Rightarrow i\gamma t + \int \left(e^{iut} - 1 - \frac{iut}{1+u^2}\right)\frac{1+u^2}{u^2}\, dG(u).$$

---

* B. V. Gnedenko [37].

From this we conclude that

$$\operatorname{Re} I_n(t) = \int (\cos ut - 1)\frac{1+u^2}{u^2}\,dG_n(u) \Rightarrow \int (\cos ut - 1)\frac{1+u^2}{u^2}\,dG(u).$$

We now deduce, in literally the same way as in the proof of Theorem 1 of § 18, that the set $\{G_n(x)\}$ is conditionally compact.

Let us take any convergent subsequence

$$G_{n_k}(x) \Rightarrow G^*(x).$$

Then

$$\int \left(e^{itu} - 1 - \frac{itu}{1+u^2}\right)\frac{1+u^2}{u^2}\,dG_{n_k}(u) \Rightarrow \int \left(e^{itu} - 1 - \frac{itu}{1+u^2}\right)\frac{1+u^2}{u^2}\,dG^*(u).$$

On the other hand,

$$i\gamma_{n_k}t + \int \left(e^{itu} - 1 - \frac{itu}{1+u^2}\right)\frac{1+u^2}{u^2}\,dG_{n_k}(u) \Rightarrow i\gamma t + \int \left(e^{itu} - 1 - \frac{itu}{1+u^2}\right)\frac{1+u^2}{u^2}\,dG(u).$$

Hence the sequence $\gamma_{n_k}$ has a limit $\gamma^*$, and by virtue of the uniqueness of the representation of an infinitely divisible law by the formula of Lévy and Khintchine we must have $\gamma^* = \gamma$, $G^*(u) \equiv G(u)$.

*Sufficiency.* From the conditions of the theorem it follows at once that

$$\log f_n(t) \to \log f(t)$$

for every $t$, that is,

$$f_n(t) \Rightarrow f(t),$$

Q.E.D.

In the sequel we need Theorem 1 in another form [37].

THEOREM 2. *For the convergence of infinitely divisible distribution functions $F_n(x)$ to a limit distribution function $F(x)$ it is necessary and sufficient that as $n \to \infty$*

$$M_n(u) \to M(u), \quad N_n(u) \to N(u), \tag{1}$$

*at the continuity points of the functions $M(u)$ and $N(u)$,*

$$\gamma_n(\tau) \to \gamma(\tau), \tag{2}$$

$$(3)\quad \lim_{\varepsilon\to 0}\overline{\lim_{n\to\infty}}\left\{\int_{-\varepsilon}^{0} u^2\,dM_n(u)+\sigma_n^2+\int_{0}^{\varepsilon} u^2 dN_n(u)\right\}$$
$$=\lim_{\varepsilon\to 0}\underline{\lim_{n\to\infty}}\left\{\int_{-\varepsilon}^{0} u^2\,dM_n(u)+\sigma_n^2+\int_{0}^{\varepsilon} u^2\,dN_n(u)\right\}=\sigma^2,$$

*where the functions $M_n(u)$, $N_n(u)$ and $M(u)$, $N(u)$ and the constants $\sigma_n$, $\gamma_n(\tau)$ and $\sigma$, $\gamma(\tau)$ are defined by* (6) *and* (9) *of* § 18 *for the distribution functions $F_n(x)$ and $F(x)$ respectively.*

*Proof.* Let $F_n \Rightarrow F$.

The necessity of the conditions (1) and (2) follows from the necessity of the condition $G_n \Rightarrow G$ of the preceding theorem, the formulas defining $M(u)$, $N(u)$, and $\gamma(\tau)$, and the theorem at the beginning of § 9.

Further, let $-\epsilon$ and $+\epsilon$ be continuity points of the functions $M_n(u)$, $M(u)$ and $N_n(u)$, $N(u)$. Then, putting

$$I_n(\varepsilon)=G_n(\varepsilon)-G_n(-\varepsilon)=\int_{-\varepsilon}^{0}\frac{u^2}{1+u^2}\,dM_n(u)+\sigma_n^2+\int_{0}^{\varepsilon}\frac{u^2}{1+u^2}\,dN_n(u),$$

$$I(\varepsilon)=G(\varepsilon)-G(-\varepsilon)=\int_{-\varepsilon}^{0}\frac{u^2}{1+u^2}\,dM(u)+\sigma^2+\int_{0}^{\varepsilon}\frac{u^2}{1+u^2}\,dN(u),$$

we have

$$I_n(\varepsilon)\to I(\varepsilon)\quad (n\to\infty). \tag{1}$$

On the basis of the relations

$$\frac{1}{1+\varepsilon^2}\int_{-\varepsilon}^{0} u^2\,dM_n(u)\leqslant\int_{-\varepsilon}^{0}\frac{u^2}{1+u^2}\,dM_n(u)\leqslant\int_{-\varepsilon}^{0} u^2\,dM_n(u),$$

$$\frac{1}{1+\varepsilon^2}\int_{0}^{\varepsilon} u^2\,dN_n(u)\leqslant\int_{0}^{\varepsilon}\frac{u^2}{1+u^2}\,dN_n(u)\leqslant\int_{0}^{\varepsilon} u^2\,dN_n(u),$$

we conclude that

$$\frac{1}{1+\varepsilon^2}\left\{\int_{-\varepsilon}^{0} u^2\,dM_n(u)+\sigma_n^2+\int_{0}^{\varepsilon} u^2 dN_n(u)\right\}\leqslant I_n(\varepsilon)$$
$$\leqslant\left\{\int_{-\varepsilon}^{0} u^2 dM_n(u)+\sigma_n^2+\int_{0}^{\varepsilon} u^2 dN_n(u)\right\}. \tag{2}$$

Finally, from (1) and (2) we deduce that *

$$\overline{\lim_{n\to\infty}}\frac{1}{1+\varepsilon^2}\left\{\int_{-\varepsilon}^{0} u^2\,dM_n(u)+\sigma_n^2+\int_{0}^{\varepsilon} u^2\,dN_n(u)\right\}\leqslant I(\varepsilon)$$
$$\leqslant\overline{\lim_{n\to\infty}}\left\{\int_{-\varepsilon}^{0} u^2\,dM_n(u)+\sigma_n^2+\int_{0}^{\varepsilon} u^2 dN_n(u)\right\}.$$

* Evidently the same inequalities hold for lower limits.

As $\epsilon$ tends to zero, both sides of the inequality above have the same limit, namely,

$$\lim_{\varepsilon \to 0} I(\varepsilon) = \sigma^2.$$

Thus the necessity of the condition (3) is proved.

Let us prove the sufficiency of the conditions. To this end we shall prove that they imply the conditions of Theorem 1.

By the theorem at the beginning of § 9, we have

$$\left. \begin{aligned} G_n(u) = \int_{-\infty}^{u} \frac{z^2}{1+z^2}\, dM_n(z) \to \int_{-\infty}^{u} \frac{z^2}{1+z^2}\, dM(z) = G(u), \\ u < 0 \, (n \to \infty) \end{aligned} \right\} \tag{3}$$

at the continuity points of $M(u)$ [and consequently also at the continuity points of $G(u)$]. From this we obtain

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} G_n(-\varepsilon) = G(-0). \tag{3'}$$

By (2), we have

$$G_n(-\varepsilon) + \frac{1}{1+\varepsilon^2} \left\{ \int_{-\varepsilon}^{0} u^2\, dM_n(u) + \sigma_n^2 + \int_{0}^{\varepsilon} u^2\, dN_n(u) \right\} \leqslant G_n(+\varepsilon)$$
$$\leqslant G_n(-\varepsilon) + \left\{ \int_{-\varepsilon}^{0} u^2\, dM_n(u) + \sigma_n^2 + \int_{0}^{\varepsilon} u^2\, dN_n(u) \right\}.$$

Hence by (3′) and the third condition of the theorem,

$$\lim_{\varepsilon \to 0} \overline{\lim_{n \to \infty}}\, G_n(+\varepsilon) = \lim_{\varepsilon \to 0} \underline{\lim_{n \to \infty}}\, G_n(+\varepsilon) = G(-0) + \sigma^2 = G(+0).$$

For every $u_1 > 0$ and $u_2 > 0$ which are continuity points of the function $G(u)$

$$\int_{u_1}^{u_2} \frac{u^2}{1+u^2}\, dN_n(u) \to \int_{u_1}^{u_2} \frac{u^2}{1+u^2}\, dN(u), \tag{4}$$

so that [$\epsilon$ is a continuity point of $N(u)$]

$$\overline{\lim_{n \to \infty}}\, G_n(u) = \lim_{\varepsilon \to 0} \overline{\lim_{n \to \infty}} \left( G_n(\varepsilon) + \int_{\varepsilon}^{u} \frac{u^2}{1+u^2}\, dN_n(u) \right)$$
$$= \lim_{\varepsilon \to 0} \underline{\lim_{n \to \infty}} \left( G_n(\varepsilon) + \int_{\varepsilon}^{u} \frac{u^2}{1+u^2}\, dN_n(u) \right) = \underline{\lim_{n \to \infty}}\, G_n(u) = G(u) \quad (u > 0). \tag{5}$$

Thus we have proved that

$$G_n(u) \to G(u) \quad (n \to \infty)$$

at all continuity points of $G(u)$.

Now we have

$$G_n(+\infty) = \int_{-\infty}^{-\varepsilon} \frac{u^2}{1+u^2} dM_n(u) + \int_{\varepsilon}^{\infty} \frac{u^2}{1+u^2} dN_n(u)$$

$$+ \int_{-\varepsilon}^{0} \frac{u^2}{1+u^2} dM_n(u) + \sigma_n^2 + \int_{0}^{\varepsilon} \frac{u^2}{1+u^2} dN_n(u). \quad (6)$$

From (3), (4), (5), and (6) we conclude that as $n \to \infty$

$$G_n(+\infty) \to G(+\infty).$$

Finally, the condition (2) of Theorem 1 follows from the conditions of Theorem 2 by the theorem at the beginning of § 9, the preceding results, and the formula defining $\gamma(\tau)$.

THEOREM 3. *For the convergence of the sequence $F_n(x)$ of infinitely divisible laws with finite variances to a limit law and the convergence of their variances to the variance of the limit law $F(x)$, it is necessary and sufficient that as $n \to \infty$*

(1) $$K_n(u) \Rightarrow K(u),$$

(2) $$\gamma_n \longrightarrow \gamma,$$

*where the functions $K_n(u)$ and $K(u)$ and the constants $\gamma_n$ and $\gamma$ are defined by Kolmogorov's formula* (10) of § 18 *for $F_n(x)$ and $F(x)$.*

*Proof*. The sufficiency of the relations (1) and (2) for the convergence is evident from Theorem 2 of § 13 and (10) of § 18.

We shall now prove the necessity of the conditions of the theorem.

Let $F_n(x) \Rightarrow F(x)$. By Theorem 1 of § 13 this is equivalent to the relation

$$f_n(t) \Rightarrow f(t)$$

and consequently to the relation

$$\log f_n(t) = i\gamma_n t + \int \{ e^{iut} - 1 - iut \} \frac{1}{u^2} dK_n(u)$$

$$\Rightarrow \log f(t) = i\gamma t + \int \{ e^{iut} - 1 - iut \} \frac{1}{u^2} dK(u). \quad (7)$$

We are now assuming that the variance of $F_n(x)$ converges to that of $F(x)$. From this and from (11) of § 18 it follows that

$$K_n(+\infty) \to K(+\infty). \quad (8)$$

For every $t$ different from zero we obviously have

$$i\gamma_n + \int \{ e^{iut} - 1 - iut \} \frac{1}{tu^2} dK_n(u)$$

$$\to i\gamma + \int \{ e^{iut} - 1 - iut \} \frac{1}{tu^2} dK(u) \quad (n \to \infty).$$

Now let $t \to 0$. The integrals then converge to zero uniformly with respect to $n$.* Hence we obtain

$$\gamma_n \to \gamma.$$

Just as in the proof of Theorem 2 of § 9, let us choose from the sequence $\{K_n(u)\}$ a subsequence $K_{n_k}(u)$ which converges to some nondecreasing function $\bar{K}(u)$ at every continuity point of the latter.

We now show that

$$\int \{ e^{iut} - 1 - iut \} \frac{1}{u^2} dK_{n_k}(u) \Rightarrow \int \{ e^{iut} - 1 - iut \} \frac{1}{u^2} d\bar{K}(u). \quad (9)$$

If the function $\bar{K}(u)$ is continuous at the points $-B$ and $+B$, then by the theorem at the beginning of § 9, for a fixed $t$ and sufficiently large $k$ we have

$$\Big| \int_{-B}^{+B} \{ e^{iut} - 1 - iut \} \frac{1}{u^2} dK_{n_k}(u) - \int_{-B}^{+B} \{ e^{iut} - 1 - iut \} \frac{1}{u^2} d\bar{K}(u) \Big| < \frac{\varepsilon}{2}. \quad (10)$$

On the other hand,

$$L_k = \Big| \int_{|u| \geqslant B} \{ e^{iut} - 1 - iut \} \frac{1}{u^2} dK_{n_k}(u) \Big| \leqslant 2|t| \int_{|u| \geqslant B} \frac{1}{|u|} dK_{n_k}(u) \leqslant \frac{2|t|}{B} \sup_k K_{n_k}(+\infty)$$

and

$$L = \Big| \int_{|u| \geqslant B} (e^{iut} - 1 - iut) \frac{1}{u^2} d\bar{K}(u) \Big| \leqslant \frac{2|t|}{B} \bar{K}(+\infty).$$

But $K_n(+\infty)$ is bounded. Therefore whatever $\epsilon > 0$ and $t$ may be, we can make

$$|L_k| < \frac{\varepsilon}{4}, \quad |L| < \frac{\varepsilon}{4} \quad (11)$$

by taking $B$ sufficiently large. (9) follows immediately from (10) and (11).

By virtue of the uniqueness of the representation by Kolmogorov's formula we conclude that

$$\bar{K}(u) = K(u).$$

---

* In fact it is easy to prove that $|e^{itu} - 1 - itu| \leq \frac{1}{2}t^2u^2$. Therefore

$$\Big| \int \{ e^{iut} - 1 - iut \} \frac{1}{tu^2} dK_n(u) \Big| \leqslant \frac{|t|}{2} \int dK_n(u).$$

and the multiplier of $|t|/2$ on the right side is bounded.

Taking into account (8), we conclude also that

$$K_{n_k}(u) \Longrightarrow K(u),$$

i.e., every weakly convergent subsequence $K_{n_k}(u)$ converges to $K(u)$. It follows that the whole sequence $K_n(u)$ converges to $K(u)$; this proves the necessity of the first condition of the theorem.

# *Part II* GENERAL LIMIT THEOREMS

## CHAPTER 4

## GENERAL LIMIT THEOREMS FOR SUMS OF INDEPENDENT SUMMANDS

### § 20. Statement of the Problem, Sums of Infinitely Divisible Summands

The most general statement of the problem of the nature of limit distribution functions for sums of independent random variables can be formulated as follows: Let $\zeta_1, \zeta_2, \ldots, \zeta_n, \ldots$ be a sequence of random variables each of which is the sum of a certain number of mutually independent random variables,

$$\zeta_n = \xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n}.$$

Suppose that for a suitable choice of real constants $A_n$ the distribution functions of the variables $\zeta_n - A_n$ converge to a certain limit. It is asked what properties this limit distribution function must possess.

The problem stated in such a general way does not, however, present any real interest if the $\xi_{nk}$ are not subject to some additional conditions, since any distribution function $F(x)$ can appear as the limit of the distribution functions of the sums

$$\xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n} - A_n. \tag{1}$$

To this end it suffices, for example, for the first term in each sum (1) to have the distribution $F(x)$, the others to be zero with probability 1, while $A_n$ is chosen to be zero.

Problems in mathematical statistics and theoretical physics which reduce in mathematical terms to the study of the limiting behavior of sums of random variables call for the same reasonable general restrictions that it is necessary to introduce in the present statement of the problem. Namely, it is to be remembered that the specific properties of the limit distribution functions should be determined by the fact that *they are the limits for sums of an increasing number of independent random variables, such that the role of a single summand becomes vanishingly small as* $n \to \infty$.

With this additional restriction (up to now formulated only in a purely qualitative manner) the problem just posed received an exhaustive solution in the papers of Bawly [1] and A. Ya. Khintchine [58]. It was proved by these two mathematicians that every distribution function which is the limit of the distribution functions of the sums (1) is infinitely divisible. This central result of the present chapter will be obtained in § 21 for the case of summands with finite variances and in § 24 in the general

case. Its proof is based on another basic proposition in the theory, namely, that under the restriction introduced above the distribution functions of the sums (1) will approach the distribution functions of the sums

$$\bar{\xi}_{n1} + \bar{\xi}_{n2} + \ldots + \bar{\xi}_{nk_n} - A_n$$

of specially constructed infinitely divisible random variables. The proof of this theorem forms the content of § 24.

In § 25 various forms of necessary and sufficient conditions for the existence of a limit distribution for the sequence of sums (1) are presented. We remark at once that these theorems can be used also to determine conditions for convergence to any given limit law.

A more precise idea of the vanishingly small role of individual summands in the formation of $\zeta_n$ is expressed by the following definition: the variables $\xi_{nk}$ are called *infinitesimal* if

$$\sup_{1 \leqslant k \leqslant k_n} \mathsf{P}\{|\xi_{nk}| \geqslant \varepsilon\} \to 0 \tag{2}$$

as $n \to \infty$ for every $\epsilon > 0$.

We see that for further derivations it is sufficient to require somewhat less. The restriction which we shall use is given by the following definition:

DEFINITION. The variables $\xi_{nk}$ are called *asymptotically constant* if it is possible to find constants $a_{nk}$ so that for every $\epsilon > 0$

$$\sup_{1 \leqslant k \leqslant k_n} \mathsf{P}\{|\xi_{nk} - a_{nk}| \geqslant \varepsilon\} \to 0 \tag{3}$$

as $n \to \infty$.

LEMMA 1. *If the $\xi_{nk}$ are asymptotically constant, then it is possible to take $a_{nk} = m_{nk}$ in (3), where $m_{nk}$ is a median of the variable $\xi_{nk}$, that is, a number such that*

$$\mathsf{P}\{\xi_{nk} \geqslant m_{nk}\} \geqslant \frac{1}{2}, \quad \mathsf{P}\{\xi_{nk} \leqslant m_{nk}\} \geqslant \frac{1}{2}.*$$

*Proof.* We remark first of all that if the probability of $\xi$ lying in some interval is greater than $\frac{1}{2}$, then every median $m\xi$ belongs to this interval.

Now let $\xi_{nk}$ be asymptotically constant. It is sufficient to prove that

$$\sup_{1 \leqslant k \leqslant k_n} |m_{nk} - a_{nk}| \to 0$$

as $n \to \infty$.

For every $\epsilon > 0$ we can find $n(\epsilon)$ such that for $n \geq n(\epsilon)$

$$\sup_{1 \leqslant k \leqslant k_n} \mathsf{P}\{|\xi_{nk} - a_{nk}| > \varepsilon\} < \frac{1}{2},$$

* There can be more than one number which satisfies the inequality; under *median* we mean any one of them.

that is,

$$\inf_{1 \leq k \leq k_n} \mathsf{P}\{|\xi_{nk} - a_{nk}| \leqslant \varepsilon\} > \frac{1}{2}.$$

Hence by the remark made at the beginning we must have

$$\sup_{1 \leq k \leq k_n} |m_{nk} - a_{nk}| \leqslant \varepsilon$$

for $n \geq n(\epsilon)$, which is what is required.

LEMMA 2. *The variables $\xi_{nk}$ are infinitesimal if and only if*

$$\sup_{1 \leqslant k \leqslant k_n} \int \frac{x^2}{1+x^2} dF_{nk}(x) \to 0 \tag{4}$$

*as $n \to \infty$.*

*Proof.* (4) follows from (2). In fact, if $0 < \epsilon < \frac{1}{2}$ and if $n$ is sufficiently large,

$$\sup_{1 \leqslant k \leqslant k_n} \int \frac{x^2}{1+x^2} dF_{nk}(x) \leqslant \sup_{1 \leqslant k \leqslant k_n} \Bigg[ \int\limits_{|x| < \varepsilon} x^2 \, dF_{nk}(x)$$

$$+ \int\limits_{|x| \geqslant \varepsilon} dF_{nk}(x) \Bigg] \leqslant \varepsilon^2 + \sup_{1 \leqslant k \leqslant k_n} \mathsf{P}\{|\xi_{nk}| \geqslant \varepsilon\} \leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Conversely, by virtue of the inequalities

$$\sup_{1 \leqslant k \leqslant k_n} \int \frac{x^2}{1+x^2} dF_{nk}(x) \geqslant \sup_{1 \leqslant k \leqslant k_n} \int\limits_{|x| \geqslant \varepsilon} \frac{x^2}{1+x^2} dF_{nk}(x)$$

$$\geqslant \frac{\varepsilon^2}{1+\varepsilon^2} \sup_{1 \leqslant k \leqslant k_n} \int\limits_{|x| \geqslant \varepsilon} dF_{nk}(x) = \frac{\varepsilon^2}{1+\varepsilon^2} \sup_{1 \leqslant k \leqslant k_n} \mathsf{P}\{|\xi_{nk}| \geqslant \varepsilon\}$$

(2) follows from (4).

We can now write down the condition for the variables $\xi_{nk}$ to be asymptotically constant as follows:

$$\sup_{1 \leqslant k \leqslant k_n} \int \frac{x^2}{1+x^2} dF_{nk}(x + m_{nk}) \to 0$$

as $n \to \infty$.

For later purposes we shall need the following:

THEOREM 1. *In order that the random variables $\xi_{nk}$ should be infinitesimal it is necessary* † *that as $n \to \infty$*

$$\sup_{1 \leqslant k \leqslant k_n} |f_{nk}(t) - 1| \Rightarrow 0.$$

† *Translator's note.* The condition is also *sufficient*, as will be needed later. The proof follows from (2) of § 14, by taking $\tau$ arbitrarily large.

*Proof.* In fact, for every $\epsilon > 0$

$$\sup_{1 \leqslant k \leqslant k_n} |f_{nk}(t) - 1| = \sup_{1 \leqslant k \leqslant k_n} \left| \int (e^{itx} - 1)\, dF_{nk}(x) \right|$$

$$\leqslant \sup_{1 \leqslant k \leqslant k_n} \int_{|x| < \epsilon} |e^{itx} - 1|\, dF_{nk}(x) + 2 \sup_{1 \leqslant k \leqslant k_n} \int_{|x| \geqslant \epsilon} dF_{nk}(x).$$

Or, taking account of the inequality $|e^{iz} - 1| \leq |z|$, valid for real $z$,

$$\sup_{1 \leqslant k \leqslant k_n} |f_{nk}(t) - 1| \leqslant \varepsilon |t| + 2 \sup_{1 \leqslant k \leqslant k_n} \mathsf{P}\{|\xi_{nk}| \geqslant \varepsilon\}.$$

Hence if the $\xi_{nk}$ are infinitesimal, we immediately obtain the required conclusion.

## § 21. Limit Distributions with Finite Variances

In this section we shall consider a double sequence

$$\xi_{n1}, \xi_{n2}, \ldots, \xi_{nk_n}$$

of random variables which are independent in each row, subject to the conditions:

$(\alpha)$
$$\sup_{1 \leqslant k \leqslant k_n} \mathsf{P}\{|\xi_{nk} - \mathsf{M}\xi_{nk}| \geqslant \varepsilon\} \to 0$$

as $n \to \infty$ for every $\epsilon > 0$.

$(\beta)$ The $\xi_{nk}$ have finite variances and

$$\mathsf{D}^2\Big(\sum_{k=1}^{k_n} \xi_{nk}\Big) = \sum_{k=1}^{k_n} \mathsf{D}^2\xi_{nk} \leqslant C,$$

where $C$ is a constant independent of $n$.

It is natural that in the study of limit laws for the sums

$$\zeta_n = \xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n} - A_n, \tag{1}$$

where the $A_n$ are suitably chosen constants, and where the variables $\xi_{nk}$ have finite variances, the case in which not only the distribution laws of the sums (1) converge to a limit but also the variances of the sums converge to the variance of the limit law, presents the greatest interest. Hence the presence and the meaning of the condition $(\beta)$ is clear.

Aside from their independent interest, the results of this section are valuable in that they give a clear idea of the proof of general theorems in the next section.

THEOREM 1.* *In order that for suitably chosen* † *constants* $A_n$ *the distribution functions of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \dots + \xi_{nk_n} - A_n \tag{2}$$

*of independent random variables subject to the conditions* ($\alpha$) *and* ($\beta$) *converge to a limit, it is necessary and sufficient that the distribution functions of certain "accompanying laws" converge. These accompanying laws are infinitely divisible and the logarithms of their characteristic functions are defined by*

$$\psi_n(t) = -iA_n t + \sum_{k=1}^{k_n} \left\{ it\mathsf{M}\xi_{nk} + \int (e^{itx} - 1)\, dF_{nk}(x + \mathsf{M}\xi_{nk}) \right\}, \tag{3}$$

*where* $F_{nk}(x)$ *is the distribution function of* $\xi_{nk}$.

*The limit laws for the two sequences of distribution functions coincide.*

*Proof.* The characteristic function of $\zeta_n$ is

$$f_n(t) = e^{-itA_n} \prod_{k=1}^{k_n} f_{nk}(t).$$

By Theorems 1 and 2 of § 13, for the convergence of the distribution functions of the sums (2) to a limit it is necessary and sufficient that as $n \to \infty$

$$f_n(t) \Rightarrow f(t), \tag{4}$$

where $f(t)$ is the characteristic function of the limit distribution function.

If we denote the distribution function of the variables $\xi'_{nk} = \xi_{nk} - \mathbf{M}\xi_{nk}$ by $F'_{nk}(x)$, then (4) can be transformed into the following equivalent form:

$$f_n(t) = e^{-iA_n t + \sum_{k=1}^{k_n} it\mathsf{M}\xi_{nk}} \prod_{k=1}^{k_n} f'_{nk}(t) \Rightarrow f(t). \tag{5}$$

Put

$$f'_{nk}(t) - 1 = \alpha_{nk}(t) = \alpha_{nk}.$$

By ($\alpha$) and Theorem 1 of § 20

$$\sup_{1 \leqslant k \leqslant k_n} |\alpha_{nk}| \Rightarrow 0.$$

Hence,‡ considering an arbitrary but fixed interval of $t$, we may suppose that from some $n$ on,

$$\sup_{1 \leqslant k \leqslant k_n} |\alpha_{nk}| < \frac{1}{2}.$$

---

* Theorem 1 is a slight modification of a theorem of Bawly [1].

† *Translator's note.* The phrase "suitably chosen" is ambiguous here. For *any given* $A_n$, the distribution functions of (2) will converge if and only if the "accompanying distribution functions," which involve the same $A_n$, converge. In some later theorems it will be shown how the $A_n$ are to be chosen.

‡ From this it also follows that from some $n$ on, $f_n(t)$ does not vanish in the interval considered, and consequently $\log f_n(t)$ is defined. This remark is needed below.

We have then the estimate

$$\begin{aligned}|\log f'_{nk}(t)-\alpha_{nk}|=|\log(1+\alpha_{nk})-\alpha_{nk}|&\leqslant\sum_{s=2}^{\infty}\frac{1}{s}|\alpha_{nk}|^s\\ &\leqslant\frac{1}{2}\sum_{s=2}^{\infty}|\alpha_{nk}|^s=\frac{1}{2}\frac{|\alpha_{nk}|^2}{1-|\alpha_{nk}|}<|\alpha_{nk}|^2.\end{aligned}\tag{6}$$

Since

$$\int x\,dF'_{nk}(x)=\mathsf{M}\xi'_{nk}=0,$$

the expression for $\alpha_{nk}$ can be written as

$$\alpha_{nk}=\int(e^{itx}-1)\,dF'_{nk}(x)=\int(e^{itx}-1-itx)\,dF'_{nk}(x).$$

But it is well known that for every real $x$

$$|e^{itx}-1-itx|\leqslant\frac{t^2x^2}{2},$$

so that

$$|\alpha_{nk}|\leqslant\frac{t^2}{2}\int x^2\,dF'_{nk}(x)=\frac{t^2}{2}\mathsf{D}^2\xi_{nk}.\tag{7}$$

Now by (5), (6), (7), and ($\beta$), we have

$$\begin{aligned}&\left|\log f_n(t)+itA_n-\sum_{k=1}^{k_n}\left\{it\mathsf{M}\xi_{nk}+\int(e^{itx}-1)\,dF'_{nk}(x)\right\}\right|\\ &\qquad=\left|\sum_{k=1}^{k_n}\left\{\log f'_{nk}(t)-\int(e^{itx}-1)\,dF'_{nk}(x)\right\}\right|\\ &\leqslant\sum_{k=1}^{k_n}|\alpha_{nk}|^2\leqslant\frac{t^2}{2}\sup_{1\leqslant k\leqslant k_n}|\alpha_{nk}|\left(\sum_{k=1}^{k_n}\mathsf{D}^2\xi_{nk}\right)\leqslant\frac{Ct^2}{2}\sup_{1\leqslant k\leqslant k_n}|\alpha_{nk}|.\end{aligned}\tag{8}$$

From this we conclude that *

$$\log f_n(t)-\psi_n(t)\Rightarrow 0.$$

Since $e^{\psi_n(t)}$ is a characteristic function and consequently does not exceed one in modulus,

$$f_n(t)-e^{\psi_n(t)}\Rightarrow 0.\tag{9}$$

But (9) is obviously equivalent to the assertion of the theorem.

Theorem 1 and Theorem 3 of § 17 imply the following:

COROLLARY. *The limit laws for the sums (2) of independent random variables, subject to the conditions ($\alpha$) and ($\beta$), are infinitely divisible.*

---

* Cf. the preceding footnote.

We shall now transform the sum (3) into the usual form for the logarithm of the characteristic function of an infinitely divisible law. For this purpose we shall consider the function

$$K_n(u) = \sum_{k=1}^{k_n} \int_{-\infty}^{u} x^2 \, dF'_{nk}(x). \tag{10}$$

Obviously, the function $K_n(u)$ is nondecreasing and satisfies the condition $K_n(-\infty) = 0$, and by $(\beta)$ $K_n(+\infty)$ is bounded.

It is easily calculated that the function $\psi_n(t)$ can be rewritten in the following form:

$$\psi_n(t) = -itA_n + it\left(\sum_{k=1}^{k_n} \mathsf{M}\xi_{nk}\right) + \int (e^{itu} - 1 - itu) \frac{1}{u^2} \, dK_n(u). \tag{11}$$

*Remark.* We remark that the variance of the sum (3) and of the infinitely divisible law defined by (3) are equal.

In fact, according to (11) of § 18 the variance of the infinitely divisible law (3) is

$$K_n(+\infty) = \sum_{k=1}^{k_n} \int x^2 \, dF_{nk}(x + \mathsf{M}\xi_{nk}) = \sum_{k=1}^{k_n} \mathsf{D}^2\xi_{nk} = \mathsf{D}^2\zeta_n.$$

The theorem just proved, together with Theorem 3 of § 19 of the preceding chapter, enables us to establish conditions for the existence of limit distribution functions for sums of independent random variables which satisfy the conditions $(\alpha)$ and $(\beta)$.

THEOREM 2.* *In order that for suitably chosen constants $A_n$ the distribution laws of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \dots + \xi_{nk_n} - A_n$$

*of independent random variables satisfying the condition $(\alpha)$ converge to a limit, and that the variances of these sums converge to the variance of the limit law, it is necessary and sufficient that there exist a nondecreasing function $K(u)$ such that*

$$K_n(u) \Rightarrow K(u).$$

*as $n \to \infty$, where $K_n(u)$ is defined by (10).*

*The constants $A_n$ may be chosen according to the formula*

$$A_n = \sum_{k=1}^{k_n} \mathsf{M}\xi_{nk} - \gamma + o(1),$$

*where $\gamma$ is any constant.*

* B. V. Gnedenko, [37], [41].

*The logarithm of the characteristic function of the limit law is given by Kolmogorov's formula* (10) *of* § 18 *with the constant* $\gamma$ *and the function* $K(u)$ *just defined.*

*Proof.* From the condition that $K_n(u) \Rightarrow K(u)$, it follows that $K_n(+\infty) = \sum_{k=1}^{k_n} \mathbf{D}^2\xi_{nk}$ is bounded; thus we find ourselves under the conditions ($\alpha$) and ($\beta$). By the preceding theorem and the remark after it, we may confine ourselves to finding conditions for the existence of a limit law for infinitely divisible laws, the logarithms of whose characteristic functions are given by (11). As we know, this limit will also be infinitely divisible. Our theorems are obtained from (10) and (11) as immediate consequences of Theorem 3 of § 19. In fact, the condition $K_n(u) \Rightarrow K(u)$ of the present theorem coincides with the condition (1) of Theorem 3 of § 19. The condition (2) of Theorem 3 of § 19 can be written as follows:

$$\gamma_n = -A_n + \sum_{k=1}^{k_n} \int x\, dF_{nk}(x) \to \gamma \qquad (n \to \infty),$$

where $\gamma$ is a constant, determined by Kolmogorov's formula for the limit law. From this relation we see that $A_n$ can be chosen as indicated in the theorem.

*Remark.* For the convergence of the distribution laws of the sums

$$\zeta_n = \xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n}$$

to a limit, and the convergence of the variances of the sums to the variance of the limit law, it is necessary and sufficient that besides the condition $K_n(u) \Rightarrow K(u)$ of Theorem 2 the following condition also be satisfied:

$$\sum_{k=1}^{k_n} \int x\, dF_{nk}(x) \to \gamma \qquad (n \to \infty).$$

It should be remarked that the preceding theorem gives not only conditions for the existence of a limit law but also conditions for the convergence of the distribution functions of the sum to any given limit law defined by the constant $\gamma$ and the function $K(u)$ according to Kolmogorov's formula.

As an application of the general theorem just proved, we consider the conditions for convergence to a normal law and to a Poisson law.

THEOREM 3. *In order that for suitably chosen constants* $A_n$ *the distribution functions of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n} - A_n$$

*of independent random variables* $\xi_{n1}, \xi_{n2}, \ldots, \xi_{nk_n}$ *converge to the normal law*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}}\, dz, \tag{12}$$

*and that the variances of* $\zeta_n$ *converge to one and that the variables* $\xi_{nk} - \mathbf{M}\xi_{nk}$ *be infinitesimal, it is necessary and sufficient that for every* $\epsilon > 0$ *the following relations hold:*

$$(1) \qquad \sum_{k=1}^{k_n} \int_{|x| \geqslant \epsilon} x^2\, dF_{nk}(x + \mathsf{M}\xi_{nk}) \to 0 \qquad (n \to \infty),$$

$$(2) \qquad \sum_{k=1}^{k_n} \int_{|x| < \epsilon} x^2\, dF_{nk}(x + \mathsf{M}\xi_{nk}) \to 1 \qquad (n \to \infty),$$

*where* $F_{nk}(x)$ *is the distribution function of* $\xi_{nk}$.

*Proof.* From the first condition of the theorem we conclude that the variables $\xi_{nk} - \mathbf{M}\xi_{nk}$ are infinitesimal. In fact, for every $\epsilon > 0$, we have

$$\sup_{1 \leqslant k \leqslant k_n} \mathsf{P}\{|\xi_{nk} - \mathsf{M}\xi_{nk}| \geqslant \varepsilon\} = \sup_{1 \leqslant k \leqslant k_n} \int_{|x| \geqslant \epsilon} dF_{nk}(x + \mathsf{M}\xi_{nk})$$

$$\leqslant \frac{1}{\epsilon^2} \sup_{1 \leqslant k \leqslant k_n} \int_{|x| \geqslant \epsilon} x^2\, dF_{nk}(x + \mathsf{M}\xi_{nk}) \to 0 \quad \text{as} \quad n \to \infty.$$

The relations (1) and (2) together prove that ($\beta$) holds.

Furthermore, we know that for the law (12)

$$\gamma = 0, \quad K(u) = \begin{cases} 0 & \text{for } u \leqslant 0, \\ 1 & \text{for } u > 0. \end{cases}$$

In our case the condition $K_n(u) \Rightarrow K(u)$ of Theorem 2 can be written as follows:

$$(1) \qquad \sum_{k=1}^{k_n} \int_{-\infty}^{u} x^2\, dF_{nk}(x + \mathsf{M}\xi_{nk}) \quad \to \quad \begin{cases} 0 & \text{for } u < 0, \\ 1 & \text{for } u > 0, \end{cases}$$

$$(2) \qquad \sum_{k=1}^{k_n} \int x^2\, dF_{nk}(x + \mathsf{M}\xi_{nk}) \to 1.$$

It is easy to see that these relations are equivalent to the conditions of the theorem.

*Particular case.* The theorem assumes an especially simple form if we suppose that for all $n$

$$\sum_{k=1}^{k_n} \mathsf{D}^2\xi_{nk} = 1.$$

*Under this condition, for the convergence of the distribution laws of the sums* (1) *to the normal law* (12) *it is necessary and sufficient that for every* $\epsilon > 0$

$$\sum_{k=1}^{k_n} \int_{|x| \geqslant \epsilon} x^2 \, dF_{nk}(x + \mathsf{M}\xi_{nk}) \to 0 \qquad (n \to \infty). \tag{13}$$

*If, moreover, for all* $k$ *and* $n$

$$\mathsf{M}\xi_{nk} = 0,$$

*then* (13) *can be rewritten as*

$$\sum_{k=1}^{k_n} \int_{|x| > \epsilon} x^2 \, dF_{nk}(x) \to 0 \qquad (n \to \infty).$$

The last relation contains the following

THEOREM 4.* *Let*

$$\xi_1, \ \xi_2, \ \ldots, \xi_n, \ \ldots$$

*be a sequence of independent random variables and let the distribution function of* $\xi_k$ *be* $F_k(x)$.

*In order that the distribution function of the normalized sums*

$$\zeta_n = \frac{\sum_{k=1}^{n} (\xi_k - \mathsf{M}\xi_k)}{B_n} \qquad \left(B_n^2 = \sum_{k=1}^{n} \mathsf{D}^2 \xi_k\right) \tag{14}$$

*converge to the normal law* (12) *and that the summands be infinitesimal, it is necessary and sufficient that Lindeberg's condition*

use $y$ =

the [illegible]

$$\frac{1}{B_n^2} \sum_{k=1}^{n} \int_{|x| > \epsilon B_n} x^2 \, dF_k(x + \mathsf{M}\xi_k) \to 0$$

*be satisfied for every* $\epsilon > 0$.

In the paper [9], S. N. Bernstein offered another derivation of this result of Lindeberg and Feller, proving that it is a consequence of the classical theorem of Lyapunov. Moreover, S. N. Bernstein proved there that Lyapunov's theorem gives a necessary and sufficient condition for the convergence of the distribution functions of the sums (14) to the normal law (12) under the additional requirement that the moments of order $2 + \delta$ of the normalized sums should converge to the corresponding moment of the normal law.

---

* Lindeberg [77], W. Feller [27].

THEOREM 5.* *In order that for suitably chosen constants $A_n$ the distribution functions of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n} - A_n \tag{15}$$

*of independent infinitesimal random variables $\xi_{nk}$ $(1 \leq k \leq k_n)$ converge as $n \to \infty$ to the Poisson law*

$$P(x) = \sum_{0 \leqslant m < x} \frac{e^{-\lambda}}{m!} \lambda^m \tag{16}$$

*and that the variance of the sums (1) converge to $\lambda$, it is necessary and sufficient that for every $\epsilon > 0$*

$$\sum_{k=1}^{k_n} \int_{|x-1| \geqslant \epsilon} x^2 \, dF_{nk}(x + \mathsf{M}\xi_{nk}) \to 0,$$

$$\sum_{k=1}^{k_n} \int_{|x-1| < \epsilon} x^2 \, dF_{nk}(x + \mathsf{M}\xi_{nk}) \to \lambda,$$

*where $F_{nk}(x)$ is the distribution function of $\xi_{nk}$.*

*The constants $A_n$ may be determined from the equation*

$$A_n = \sum_{k=1}^{k_n} \mathsf{M}\xi_{nk} - \lambda.$$

*Proof.* Theorem 5 follows readily from Theorem 2 and the fact that for the law (16) $\gamma = \lambda$, $K(u) = 0$ for $u \leq 1$ and $K(u) = \lambda$ for $u > 1$.

Just as in the case of the normal law, the conditions of the theorem assume the most simple form if we suppose that for each $n$ the variance of the sum $\xi_n$ is equal to the variance of the limit law, i.e., if

$$\sum_{k=1}^{k_n} \mathsf{D}^2 \xi_{nk} = \lambda.$$

Under this additional assumption, for the convergence of the distribution functions of the sums (15) to the Poisson law (16) it is necessary and sufficient that for every $\epsilon > 0$ the following relation be satisfied:

$$\sum_{k=1}^{k_n} \int_{|x-1| > \epsilon} x^2 \, dF_{nk}(x + \mathsf{M}\xi_{nk}) \to 0 \qquad (n \to \infty).$$

* B. V. Gnedenko [41].

## § 22. Law of Large Numbers

We shall begin the study of the limiting behavior of sums of independent variables in the general case with the law of large numbers. The conditions found here form the basis of proof of subsequent theorems.

Definition 1. The sequence of random variables

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

*converges in probability* to the random variable $\xi$, if for every $\epsilon > 0$

$$\mathsf{P}\{|\xi_n - \xi| \geqslant \varepsilon\} \to 0 \quad \text{as} \quad n \to \infty.$$

In what follows we shall denote convergence in probability by the symbol $\xrightarrow{P}$. The preceding relation can then be written as

$$\xi_n \xrightarrow{P} \xi \qquad (n \to \infty).$$

Definition 2. The sequence $\{\xi_n\}$ is called *stable* if there exists a sequence of constants $\{A_n\}$ such that as $n \to \infty$

$$\xi_n - A_n \xrightarrow{P} 0. \tag{1}$$

Repeating the argument carried out in § 20 (Lemma 1), we see that if the sequence $\{\xi_n\}$ is stable, then the constants $A_n$ in (1) may be taken to be the medians $m_n$. In other words, it follows from (1) that as $n \to \infty$

$$\xi_n - m_n \xrightarrow{P} 0.$$

Definition 3. The double sequence of random variables

$$\xi_{n1}, \xi_{n2}, \ldots, \xi_{nk_n} \qquad (n = 1, 2, 3, \ldots) \tag{2}$$

*obeys the law of large numbers* if the sequence of sums

$$\zeta_n = \xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n}$$

is stable.

Our next problem consists in the determination of the most general conditions under which the law of large numbers holds.

Theorem.* *In order that the double sequence* (2) *obey the law of large numbers, it is necessary and sufficient that as* $n \to \infty$

$$\left.\begin{aligned} &1)\ \sum_{k=1}^{k_n} \int_{|x|>1} dF_{nk}(x + m_{nk}) \to 0, \\ &2)\ \sum_{k=1}^{k_n} \int_{|x|\leqslant 1} x^2\, dF_{nk}(x + m_{nk}) \to 0. \end{aligned}\right\} \tag{3}$$

---

* See [63], [28], [41].

Proof → p 106

*Proof.* The sufficiency of the conditions of the theorem can be derived in a completely elementary way. For this purpose we introduce the notations

$$\xi'_{nk} = \xi_{nk} - m_{nk},$$

$$F'_{nk}(x) = \mathsf{P}\{\xi'_{nk} < x\} = F_{nk}(x + m_{nk}).$$

Furthermore, we put

$$\xi''_{nk} = \begin{cases} \xi'_{nk}, & \text{if } |\xi'_{nk}| \leqslant 1, \\ 0, & \text{if } |\xi'_{nk}| > 1 \end{cases}$$

and

$$A_n = \sum_{k=1}^{k_n} (m_{nk} + \mathsf{M}\,\xi''_{nk}),$$

where

$$\mathsf{M}\,\xi''_{nk} = \int_{|x| \leqslant 1} x\,dF'_{nk}(x).$$

Let

$$\zeta'_n = \sum_{k=1}^{k_n} \xi'_{nk}, \quad \zeta''_n = \sum_{k=1}^{k_n} \xi''_{nk}$$

and let $B_n$ be the event that $\zeta'_n = \zeta''_n$. If $\overline{B}_n$ is the event complementary to $B_n$, then it is evident that

$$\mathsf{P}\{\overline{B}_n\} \leqslant \sum_{k=1}^{k_n} \mathsf{P}\{|\xi'_{nk}| > 1\} = \sum_{k=1}^{k_n} \int_{|x|>1} dF'_{nk}(x). \tag{4}$$

Obviously, for every $\epsilon > 0$

$$\mathsf{P}\{|\zeta_n - A_n| \geqslant \varepsilon\} = \mathsf{P}\{B_n\}\,\mathsf{P}\{|\zeta_n - A_n| \geqslant \varepsilon \mid B_n\} + \mathsf{P}\{\overline{B}_n\}\,\mathsf{P}\{|\zeta_n - A_n| \geqslant \varepsilon \mid \overline{B}_n\}. \tag{5}$$

Since

$$\mathsf{P}\{|\zeta_n - A_n| \geqslant \varepsilon \mid B_n\}\,\mathsf{P}\{B_n\} \leqslant \mathsf{P}\{|\zeta''_n - \mathsf{M}\,\zeta''_n| \geqslant \varepsilon\},$$

according to Chebyshev's inequality

$$\mathsf{P}\{|\zeta_n - A_n| \geqslant \varepsilon \mid B_n\}\,\mathsf{P}\{B_n\} \leqslant \frac{1}{\varepsilon^2}\mathsf{D}^2\zeta''_n$$

$$= \frac{1}{\varepsilon^2}\sum_{k=1}^{k_n} \mathsf{D}^2\xi''_{nk} \leqslant \frac{1}{\varepsilon^2}\sum_{k=1}^{k_n} \mathsf{M}\,\xi''^2_{nk} = \frac{1}{\varepsilon^2}\sum_{k=1}^{k_n} \int_{|x|<1} x^2\,dF'_{nk}(x). \tag{6}$$

From (4), (5), and (6) taken together, we find that for every $\epsilon > 0$

$$\mathsf{P}\{|\zeta_n - A_n| \geqslant \varepsilon\} \leqslant \frac{1}{\varepsilon^2}\sum_{k=1}^{k_n} \int_{|x| \leqslant 1} x^2\,dF'_{nk}(x) + \sum_{k=1}^{k_n} \int_{|x|>1} dF'_{nk}(x),$$

which proves the sufficiency of the conditions of the theorem.

To prove the necessity of the conditions of the theorem, we make use of the apparatus of characteristic functions, for elementary methods lead to very cumbersome arguments.*

If the law of large numbers holds, then for some $A_n$

$$\zeta_n - A_n \xrightarrow{P} 0$$

as $n \to \infty$. In other words, as $n \to \infty$ the distribution function of the sum $\zeta_n - A_n$ converges to the unitary law $\epsilon(x)$ or, in terms of characteristic functions,

$$e^{-iA_n t} \prod_{k=1}^{k_n} f_{nk}(t) \Longrightarrow 1 \quad (n \to \infty).$$

From this we conclude that as $n \to \infty$

$$\prod_{k=1}^{k_n} |f_{nk}(t)| \Longrightarrow 1. \tag{7}$$

From (7) we deduce in particular that as $n \to \infty$

$$\sup_{1 \leqslant k \leqslant k_n} (1 - |f_{nk}(t)|) \Longrightarrow 0. \tag{8}$$

According to Theorem 3 of § 14 † this implies the asymptotic constancy of the variables $\xi_{nk}$. Thus, if the sums $\zeta_n$ of independent random variables are stable, the summands are asymptotically constant.

From (7) and the inequality

$$-\log(1-a) \geqslant a \qquad (0 \leqslant a < 1),$$

we easily find that as $n \to \infty$

$$\sum_{k=1}^{k_n} (1 - |f_{nk}(t)|) \Longrightarrow 0. \tag{9}$$

We shall first prove the necessity of the condition of the theorem in the particular case that all the summands $\xi_{nk}$ are symmetrical. A little later we shall reduce the general case to this particular case.

Let $f(t)$ be the characteristic function of the symmetrical random variable $\xi$. Hence $m\xi = 0$ and

$$f(t) = \int \cos tx \, dF(x).$$

We have

$$\frac{1}{2}\int_{-1}^{1} (1 - f(t))\, dt = \int \left(1 - \frac{\sin x}{x}\right) dF(x).$$

* See [63].

† *Translator's note.* This is not sufficient. We need the amended Theorem 1 of § 20, trivially modified for asymptotic constancy.

Now for $|x| > 1$

$$1 - \frac{\sin x}{x} \geqslant \frac{1}{10},$$

and for $|x| \leq 1$

$$1 - \frac{\sin x}{x} \geqslant \frac{x^2}{8}.$$

Thus

$$\int_{-1}^{1} (1 - f(t))\, dt \geqslant \frac{1}{4} \int_{|x| \leqslant 1} x^2\, dF(x) + \frac{1}{5} \int_{|x| > 1} dF(x)$$

and so

$$\int_{-1}^{1} \sum_{k=1}^{k_n} (1 - f_{nk}(t))\, dt = \sum_{k=1}^{k_n} \int_{-1}^{1} (1 - f_{nk}(t))\, dt$$

$$\geqslant \frac{1}{4} \sum_{k=1}^{k_n} \int_{|x| \leqslant 1} x^2\, dF_{nk}(x) + \frac{1}{5} \sum_{k=1}^{k_n} \int_{|x| > 1} dF_{nk}(x) \geqslant 0.$$

From this and (9) follows the necessity of the conditions of the theorem. [Remember that in the symmetrical case $m_{nk} = 0$ and $f_{nk}(t)$ is real. By (8), from some $n$ on, $|f_{nk}(t)| = f_{nk}(t)$ throughout the interval $|t| \leq 1$.]

We turn to the proof of the necessity of the conditions of the theorem in the general case. Consider the random variables $\eta_{nk}$, independent among themselves and of all $\xi_{nk}$ for each $n$, and such that $\xi_{nk}$ and $\eta_{nk}$ have the same distribution.

Put

$$\xi^*_{nk} = \xi_{nk} - \eta_{nk}$$

and

$$\zeta^*_n = \sum_{k=1}^{k_n} \xi^*_{nk}.$$

The variables $\xi^*_{nk}$ are symmetrical. Their characteristic functions are equal to $|f_{nk}(t)|^2$.

Consequently, by (7), $\zeta^*_n$ converges in probability to zero.

By what we have proved this means that as $n \to \infty$

$$\sum_{k=1}^{k_n} \int_{|x| \leqslant 1} x^2\, dF^*_{nk}(x) \to 0,$$

$$\sum_{k=1}^{k_n} \int_{|x| > 1} dF^*_{nk}(x) \to 0.$$

Hence we obtain (3) with the help of the following lemma.

LEMMA. *Let the random variables $\xi$ and $\eta$ be independent and identically distributed, and the function $s_1(x)$ be defined to be*

$$s_1(x) = \begin{cases} x^2 & \text{for } |x| \leqslant 1, \\ 1 & \text{for } |x| > 1. \end{cases}$$

*Then*

$$\mathsf{M}s_1(\xi - \eta) \geqslant \frac{1}{2} \mathsf{M}s_1(\xi - m),$$

*where $m$ is the median of $\xi$.*

*Proof.* Without loss of generality we may suppose that $m = 0$, since otherwise we may consider the random variables $\xi - m$ and $\eta - m$. Obviously, if $x$ and $y$ are of opposite signs, $s_1(x - y) \geq s_1(x)$.

Therefore

$$\begin{aligned} \mathsf{M}s_1(\xi - \eta) &= \int\int s_1(x - y)\, dF(x)\, dF(y) \\ &\geqslant \int\limits_{\{x \geqslant 0,\ y \leqslant 0\}}\int + \int\limits_{\{x < 0,\ y \geqslant 0\}}\int s_1(x - y)\, dF(x)\, dF(y) \\ &\geqslant \int\limits_{y \leqslant 0} dF(y) \int\limits_{x \geqslant 0} s_1(x)\, dF(x) + \int\limits_{y \geqslant 0} dF(y) \int\limits_{x < 0} s_1(x)\, dF(x) \\ &\geqslant \frac{1}{2} \int s_1(x)\, dF(x) = \frac{1}{2} \mathsf{M}s_1(\xi). \end{aligned}$$

*Remark.* Obviously, with the help of the function $s_1(x)$ the conditions of the theorem can be written as follows: as $n \to \infty$

$$\sum_{k=1}^{k_n} \mathsf{M}s_1(\xi_{nk} - m_{nk}) \to 0.$$

Since

$$\frac{x^2}{1 + x^2} \leqslant s_1(x) \leqslant 2 \frac{x^2}{1 + x^2},$$

it is evident that the conditions of the theorem are also equivalent to the condition that as $n \to \infty$

$$\sum_{k=1}^{k_n} \mathsf{M} \frac{(\xi_{nk} - m_{nk})^2}{1 + (\xi_{nk} - m_{nk})^2} \to 0. \tag{10}$$

## § 23. Two Auxiliary Theorems

We shall now turn to the general problem of determining the limit laws for sums of independent random variables. We confine ourselves to considering a double sequence

$$\xi_{n1}, \xi_{n2}, \ldots, \xi_{nk_n}$$

of random variables which are independent in each row and asymptotically constant.

We begin the investigation of this problem with the proof of two theorems which show that even in the general case, where we can no longer require the existence of variances for the summands, there is a valid inequality analogous to the condition ($\beta$) of § 21. This circumstance enables us to retain the idea of proof of the theorems of § 21 and to extend them to cover the general case.

THEOREM 1.* *If for some suitably chosen constants* $A_n$ *the distribution functions of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n} - A_n$$

*of independent random variables* $\xi_{nk}$ *converge to a limit, then there exists a constant* $C < \infty$ *such that*

$$\sum_{k=1}^{k_n} \int \frac{x^2}{1+x^2}\, dF_{nk}(x + m_k) < C. \tag{1}$$

*Proof.* This theorem, which is basic for all that follows, is a simple consequence of the results of the preceding section.

By hypothesis, the distribution functions of the sums $\zeta_n$ converge to a limit as $n \to \infty$. Hence it follows easily that for *every* sequence of constants

$$\alpha_n \to 0$$

we have

$$\alpha_n \zeta_n \xrightarrow{P} 0.$$

But for $0 < \alpha < 1$,

$$\mathsf{M}\frac{\alpha^2\xi^2}{1+\alpha^2\xi^2} = \alpha^2 \mathsf{M}\frac{\xi^2}{1+\alpha^2\xi^2} \geqslant \alpha^2 \mathsf{M}\frac{\xi^2}{1+\xi^2}.$$

Hence for sufficiently large $n$ ($\xi'_{nk} = \xi_{nk} - m_{nk}$)

$$\sum_{k=1}^{k_n} \mathsf{M}\frac{\alpha_n^2 \xi'^2_{nk}}{1+\alpha_n^2\xi'^2_{nk}} \geqslant \alpha_n^2 \sum_{k=1}^{k_n} \mathsf{M}\frac{\xi'^2_{nk}}{1+\xi'^2_{nk}}$$

and, consequently, as $n \to \infty$

$$\alpha_n^2 \sum_{k=1}^{k_n} \int \frac{x^2}{1+x^2}\, dF_{nk}(x+m_{nk}) \to 0.$$

If the sums (1) were not bounded, then the preceding relation could not hold for every sequence $\alpha_n \to 0$.

*Remark.* It is possible to give another proof of Theorem 1 without making use of the results of the preceding section. In this respect it is interesting that further results, among which the law of large numbers is also counted, can thus be obtained by a single method without using the results of § 22 (see B. V. Gnedenko [41]).

* For another proof see B. V. Gnedenko [41].

THEOREM 2.* *If for some suitably chosen constants $A_n$ the distribution functions of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \dots + \xi_{nk_n} - A_n$$

*of independent infinitesimal random variables converge to a limit as $n \to \infty$, then there exists a constant $C$ such that*

$$\sum_{k=1}^{k_n} \int \frac{x^2}{1+x^2} dF_{nk}(x + \alpha_{nk}) < C,$$

*where*

$$\alpha_{nk} = \int_{|x|<\tau} x\, dF_{nk}(x),$$

*and $\tau$ is any positive constant.*

*Proof.* We note first that since the variables $\xi_{nk}$ are infinitesimal the following relations hold†

$$\sup_{1\leqslant k\leqslant k_n} |m_{nk}| \to 0 \qquad (n \to \infty),$$

$$\sup_{1\leqslant k\leqslant k_n} |\alpha_{nk}| \to 0 \qquad (n \to \infty).$$

Because of the elementary inequality $(a+b)^2 \leq 2(a^2+b^2)$, we conclude that for sufficiently large $n$

$$\begin{aligned}\int \frac{x^2}{1+x^2} dF_{nk}(x+\alpha_{nk}) &= \int \frac{(x+m_{nk}-\alpha_{nk})^2}{1+(x+m_{nk}-\alpha_{nk})^2} dF_{nk}(x+m_{nk}) \\ &\leqslant \int \frac{2x^2\, dF_{nk}(x+m_{nk})}{1+(x+m_{nk}-\alpha_{nk})^2} + 2(m_{nk}-\alpha_{nk})^2 \\ &\leqslant c \int \frac{x^2}{1+x^2} dF_{nk}(x+m_{nk}) + 2(m_{nk}-\alpha_{nk})^2.\end{aligned}$$

Now we estimate the difference $(m_{nk} - \alpha_{nk})^2$. We have, for sufficiently large $n$,

$$\begin{aligned}(\alpha_{nk}-m_{nk})^2 &= \Bigl(\int_{|x|<\tau} (x-m_{nk})\, dF_{nk}(x) - \int_{|x|\geqslant\tau} m_{nk}\, dF_{nk}(x)\Bigr)^2 \\ &\leqslant 2\Bigl(\int_{|x+m_{nk}|<\tau} |x|\, dF_{nk}(x+m_{nk})\Bigr)^2 \\ &\quad + 2\Bigl(\int_{|x+m_{nk}|\geqslant\tau} m_{nk}\, dF_{nk}(x+m_{nk})\Bigr)^2 \leqslant \qquad \text{(cont'd)}\end{aligned}$$

---

* B. V. Gnedenko [41].

† The second relation is derived as follows:

$$|\alpha_{nk}| = \Bigl|\int_{|x|<\tau} x\, dF_{nk}(x)\Bigr| \leqslant \int_{|x|\leqslant\epsilon} |x|\, dF_{nk}(x) + \tau P\{|\xi_{nk}| \geqslant \epsilon\}$$
$$\leqslant \epsilon + \tau P\{|\xi_{nk}| \geqslant \epsilon\}.$$

Choosing $\epsilon > 0$ sufficiently small, and $n$ sufficiently large, we can make $\sup\limits_{1\leq k\leq k_n} |\alpha_{nk}|$ as small as we wish.

$$\leqslant 2\Big(\int\limits_{|x|<2\tau} |x|\, dF_{nk}(x+m_{nk})\Big)^2$$
$$+2m_{nk}^2\Big(\int\limits_{|x|>\frac{\tau}{2}} dF_{nk}(x+m_{nk})\Big)^2.$$

Finally, applying the inequality of Cauchy and Bunyakovski, we find that

$$(\alpha_{nk}-m_{nk})^2 \leqslant 2\int\limits_{|x|<2\tau} x^2 dF_{nk}(x+m_{nk}) + 2m_{nk}^2 \int\limits_{|x|>\frac{\tau}{2}} dF_{nk}(x+m_{nk}).$$

From the inequalities obtained and from Theorem 1 follows the assertion of Theorem 2.

## § 24. The General Form of the Limit Theorems. The Accompanying Infinitely Divisible Laws

THEOREM 1.* *In order that for some suitably chosen constants $A_n$ the distribution functions of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \dots + \xi_{nk_n} - A_n \tag{1}$$

*of independent infinitesimal random variables converge to a limit, it is necessary and sufficient that the infinitely divisible laws,† the logarithms of whose characteristic functions are given by the formula*

$$\psi_n(t) = -iA_n t + \sum_{k=1}^{k_n}\Big\{it\alpha_{nk} + \int (e^{itx}-1)\, dF_{nk}(x+\alpha_{nk})\Big\}, \tag{2}$$

*converge. Here*

$$\alpha_{nk} = \int\limits_{|x|<\tau} x\, dF_{nk}(x), \tag{3}$$

*and $\tau > 0$ is a constant. The limit distribution functions for the two sequences coincide.*

*Proof.* In order that the distribution functions of the sums converge to a limit, it is necessary and sufficient that

$$f_n(t) = e^{-itA_n}\prod_{k=1}^{k_n} f_{nk}(t) \Rightarrow f(t) \qquad (n\to\infty), \tag{4}$$

where $f(t)$ is the characteristic function of the limit law and $f_n(t)$ is the characteristic function of the sum (1).

We introduce the notation

$$F'_{nk}(x) = F_{nk}(x+\alpha_{nk}),$$

where $\alpha_{nk}$ is defined by (3).

---

* B. V. Gnedenko [36], [37], [41].

† We shall again call them the *accompanying* laws.

Then (4) can be written in the following equivalent form:

$$e^{-itA_n+it\sum_{k=1}^{k_n}\alpha_{nk}}\prod_{k=1}^{k_n} f'_{nk}(t) \Rightarrow f(t) \qquad (n\to\infty).$$

Set

$$f'_{nk}(t)-1=\beta_{nk};$$

since the variables $\xi_{nk}$ are infinitesimal,

$$\sup_{1\leqslant k\leqslant k_n}|\beta_{nk}|\to 0 \qquad (n\to\infty), \quad (5)$$

we can make use of the expansion of the logarithm in a series:

$$\log f'_{nk}(t)=\log(1+\beta_{nk})=\beta_{nk}-\frac{1}{2}\beta_{nk}^2+\frac{1}{3}\beta_{nk}^3-\dots$$

Hence, for sufficiently large $n$, we find

$$\left|\sum_{k=1}^{k_n}\log f'_{nk}(t)-\sum_{k=1}^{k_n}\beta_{nk}\right|\leqslant\sum_{k=1}^{k_n}\sum_{s=2}^{\infty}\frac{1}{s}|\beta_{nk}|^s$$

$$\leqslant\frac{1}{2}\sum_{k=1}^{k_n}\frac{|\beta_{nk}|^2}{1-|\beta_{nk}|}\leqslant\sup_{1\leqslant k\leqslant k_n}|\beta_{nk}|\sum_{k=1}^{k_n}|\beta_{nk}|. \quad (6)$$

Now

$$|\beta_{nk}|=\left|\int_{|x|<\tau}(e^{itx}-1-itx)\,dF'_{nk}(x)+\int_{|x|\geqslant\tau}(e^{itx}-1)\,dF'_{nk}(x)\right.$$

$$\left.+it\int_{|x|<\tau}x\,dF'_{nk}(x)\right|\leqslant\frac{1}{2}|t|^2\int_{|x|<\tau}x^2\,dF'_{nk}(x)$$

$$+2\int_{|x|\geqslant\tau}dF'_{nk}(x)+|t|\left|\int_{|x|<\tau}x\,dF'_{nk}(x)\right|. \quad (7)$$

Now we estimate

$$\int_{|x|<\tau}x\,dF'_{nk}(x).$$

We have for sufficiently large $n$ $\left(\text{such that } |\alpha_{nk}|<\frac{\tau}{2}\right)$

$$\left|\int_{|x|<\tau}x\,dF'_{nk}(x)-\int_{|x+\alpha_{nk}|<\tau}x\,dF'_{nk}(x)\right|\leqslant\int_{\frac{\tau}{2}<|x|<\frac{3\tau}{2}}|x|\,dF'_{nk}(x)$$

$$\leqslant\frac{3\tau}{2}\int_{|x|>\frac{\tau}{2}}dF'_{nk}(x)$$

and

$$\Big|\int\limits_{|x+\alpha_{nk}|<\tau} x\, dF'_{nk}(x)\Big| = \Big|\int\limits_{|x|<\tau} (x-\alpha_{nk})\, dF_{nk}(x)\Big|$$

$$= \Big|\alpha_{nk}\int\limits_{|x|>\tau} dF_{nk}(x)\Big| \leqslant \frac{\tau}{2}\int\limits_{|x|>\frac{\tau}{2}} dF'_{nk}(x).$$

Therefore

$$\Big|\int\limits_{|x|<\tau} x\, dF'_{nk}(x)\Big| \leqslant 2\tau \int\limits_{|x|>\frac{\tau}{2}} dF'_{nk}(x). \tag{8}$$

To prove the necessity of the conditions of the theorem we notice that by Theorem 2 of § 23 if the distribution functions of the sums (1) converge to a limit, then

$$\sum_{k=1}^{k_n}\int \frac{x^2}{1+x^2}\, dF'_{nk}(x) < C. \tag{9}$$

We have, consequently,

$$\sum_{k=1}^{k_n}\int\limits_{|x|<\tau} x^2\, dF'_{nk}(x) \leqslant (1+\tau^2)\sum_{k=1}^{k_n}\int\limits_{|x|<\tau}\frac{x^2}{1+x^2}\, dF'_{nk}(x) \leqslant (1+\tau^2)\, C,$$

$$\sum_{k=1}^{k_n}\int\limits_{|x|\geqslant\tau} dF'_{nk}(x) \leqslant \frac{1+\tau^2}{\tau^2}\sum_{k=1}^{k_n}\int\limits_{|x|\geqslant\tau}\frac{x^2}{1+x^2}\, dF'_{nk}(x) \leqslant \frac{1+\tau^2}{\tau^2}\, C.$$

Therefore, by virtue of (7) and (8),

$$\sum_{k=1}^{k_n}|\beta_{nk}| \leqslant \left\{\frac{1+\tau^2}{2}|t|^2 + 2\,\frac{(1+\tau^2)}{\tau^2} + 2\,\frac{|t|\,(4+\tau^2)}{\tau}\right\} C,$$

where $C$ is a constant. Thus it follows from (5) and (6) that as $n\to\infty$

$$|\log f_n(t) - \psi_n(t)| = \Big|\Big[-itA_n + it\sum_{k=1}^{k_n}\alpha_{nk} + \sum_{k=1}^{k_n}\log f'_{nk}(t)\Big]$$

$$-\Big[-itA_n + it\sum_{k=1}^{k_n}\alpha_{nk} + \sum_{k=1}^{k_n}\beta_{nk}\Big]\Big|$$

$$= \Big|\sum_{k=1}^{k_n}\log f'_{nk}(t) - \sum_{k=1}^{k_n}\beta_{nk}\Big| \Rightarrow 0.$$

Since $e^{\psi_n(t)}$ is a characteristic function and hence cannot exceed one in modulus, we have

$$f_n(t) - e^{\psi_n(t)} \Longrightarrow 0, \tag{10}$$

which proves the necessity of the conditions of the theorem.

To prove the sufficiency of the conditions of the theorem we have also to estimate the sum

$$\sum_{k=1}^{k_n} |\beta_{nk}|.$$

For this purpose we note that for the infinitely divisible laws defined by (2), we should put

$$G_n(u) = \sum_{k=1}^{k_n} \int_{-\infty}^{u} \frac{x^2}{1+x^2} dF'_{nk}(x)$$

in the formula of Lévy and Khintchine.

Hence, if the infinitely divisible laws defined by (2) converge to a limit, then by Theorem 1 of § 19

$$\int dG_n(u) = \sum_{k=1}^{k_n} \int \frac{x^2}{1+x^2} dF'_{nk}(x) \to \int dG(u) \qquad (n \to \infty),$$

where $G(u)$ is the monotone function defined by the formula of Lévy and Khintchine for the limit law.

Thus we also have the relation (9) if the functions (2) converge to a limit. Therefore from (5) and (6) we again conclude that (10) holds. The proof of the theorem is thereby completed.

The theorem just proved is of considerable interest, since it permits us to replace the investigation of sums

$$\zeta_n = \xi_{n1} + \xi_{n2} + \dots + \xi_{nk_n} - A_n$$

of infinitesimal random variables $\xi_{nk}$ with, generally speaking, arbitrary distribution functions $F_{nk}(x)$ by an investigation of sums

$$\bar{\zeta}_n = \bar{\xi}_{n1} + \bar{\xi}_{n2} + \dots + \bar{\xi}_{nk_n} - A_n$$

of infinitely divisible variables $\bar{\xi}_{nk}$. This circumstance, as already stated in § 20, is made the basis for the exposition of theorems concerning the limit distributions for sums of independent variables.

As a first consequence of the theorem, we cite the following fundamental result, obtained by A. Ya. Khintchine [58].

THEOREM 2. *In order that $F(x)$ be the limit distribution function of sums* (1) *of infinitesimal random variables which are independent in each row, it is necessary and sufficient that $F(x)$ be infinitely divisible.*

*Proof.* From the preceding theorem we know that the limit distribution function for the sums (1) is simultaneously the limit of infinitely divisible laws and so according to Theorem 3 of § 17 is itself infinitely divisible. The converse proposition, that every infinitely divisible law is the limit law for sums of infinitesimal variables, follows readily from the definition of infinitely divisible laws.

*Thus we have proved that the class of limit laws for the sums* (1) *of independent infinitesimal random variables coincides with the class of infinitely divisible laws.*

If the $\xi_{nk}$ $(1 \leq k \leq k_n,\ n = 1, 2, \ldots)$ are asymptotically constant random variables, then the variables $\xi_{nk} - m_{nk}$ are infinitesimal. Therefore the class of limit laws for sums (1) of asymptotically constant variables (not only of infinitesimal variables) coincides with the class of infinitely divisible laws.

## § 25. Necessary and Sufficient Conditions for Convergence

Theorem 1 of § 24 enables us to find conditions for the existence of a limit distribution function for the sums (1) of § 24.

THEOREM 1. *In order that for some suitably chosen constants $A_n$ the distributions of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n} - A_n \tag{1}$$

*of independent infinitesimal random variables converge to a limit, it is necessary and sufficient that there exist nondecreasing functions*

$$M(u)\ (M(-\infty) = 0) \text{ and } N(u)\ (N(+\infty) = 0),$$

*defined in the intervals $(-\infty, 0)$ and $(0, +\infty)$ respectively, and a constant $\sigma \geq 0$, such that*

1) *At every continuity point of $M(u)$ and $N(u)$*

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} F_{nk}(u) = M(u) \qquad (u < 0),$$

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} (F_{nk}(u) - 1) = N(u), \qquad (u > 0),$$

2)
$$\lim_{\varepsilon\to 0} \varliminf_{n\to\infty} \sum_{k=1}^{k_n} \left\{ \int_{|x|<\varepsilon} x^2\, dF_{nk}(x) - \left( \int_{|x|<\varepsilon} x\, dF_{nk}(x) \right)^2 \right\}$$

$$= \lim_{\varepsilon\to 0} \varlimsup_{n\to\infty} \sum_{k=1}^{k_n} \left\{ \int_{|x|<\varepsilon} x^2 dF_{nk}(x) - \left( \int_{|x|<\varepsilon} x\, dF_{nk}(x) \right)^2 \right\} = \sigma^2.$$

*The constants $A_n$ may be chosen according to the formula*

$$A_n = \sum_{k=1}^{k_n} \int_{|x|<\tau} x\, dF_{nk}(x) - \gamma(\tau),$$

*where $\gamma(\tau)$* is any constant and $-\tau$ and $+\tau$ are continuity points of $M(u)$ and $N(u)$.*

*The logarithm of the characteristic function of the limit law is defined by the formula (7) of § 18 with the functions $M(u)$, $N(u)$ and the above constants $\sigma$, $\gamma(\tau)$.*

*Proof.* The theorem formulated above is a consequence of Theorems 2 of § 19 and 1 of § 24. In fact, by Theorem 1 of § 24 we may confine ourselves to the investigation of conditions of convergence of infinitely divisible laws defined by (2), § 24. For these laws we should take in (8) of § 18:

$$M_n(u) = \sum_{k=1}^{k_n} \int_{-\infty}^{u} dF_{nk}(x+\alpha_{nk}) \qquad \text{for} \quad u<0,$$

$$N_n(u) = -\sum_{k=1}^{k_n} \int_{u}^{\infty} dF_{nk}(x+\alpha_{nk}) \quad \text{for} \quad u>0,$$

$$\sigma_n = 0 \quad \text{and} \quad \gamma_n(\tau) = -A_n + \sum_{k=1}^{k_n} \int_{|x|<\tau} x\, dF_{nk}(x) + \sum_{k=1}^{k_n} \int_{|x|<\tau} x\, dF'_{nk}(x).$$

According to Theorem 2 of § 19 for the convergence of the distribution laws (2) of § 24 it is necessary and sufficient that as $n \to \infty$

$$1') \quad \sum_{k=1}^{k_n} \int_{-\infty}^{u} dF_{nk}(x+\alpha_{nk}) \to M(u) \quad (u<0);$$

$$-\sum_{k=1}^{k_n} \int_{u}^{\infty} dF_{nk}(x+\alpha_{nk}) \to N(u) \quad (u>0);$$

$$2') \lim_{\varepsilon\to 0} \underline{\lim}_{n\to\infty} \sum_{k=1}^{k_n} \left[ \int_{-\varepsilon}^{0} x^2\, dF_{nk}(x) + \int_{0}^{\varepsilon} x^2\, dF_{nk}(x) \right]$$

$$= \lim_{\varepsilon\to 0} \underline{\lim}_{n\to\infty} \sum_{k=1}^{k_n} \int_{-\varepsilon}^{+\varepsilon} x^2\, dF_{nk}(x) = \lim_{\varepsilon\to 0} \overline{\lim}_{n\to\infty} \sum_{k=1}^{k_n} \int_{-\varepsilon}^{+\varepsilon} x^2\, dF_{nk}(x) = \sigma^2,$$

$$3') - A_n + \sum_{k=1}^{k_n} \int_{|x|<\tau} x\, dF_{nk}(x) + \sum_{k=1}^{k_n} \int_{|x|<\tau} x\, dF'_{nk}(x) \to \gamma(\tau).$$

---

* *Translator's note.* The argument $\tau$ apparently serves to recall formula (8) of § 18.

From 1′), 2′), 3′) we deduce first of all that $A_n$ may be chosen as indicated in Theorem 1. For this purpose we shall prove that

$$\sum_{k=1}^{k_n} \int_{|x|<\tau} x\, dF'_{nk}(x) \to 0 \quad (n \to \infty). \tag{2}$$

We have

$$\sum_{k=1}^{k_n} \int_{|x|<\tau} x\, dF'_{nk}(x) = \sum_{k=1}^{k_n} \int_{|x-\alpha_{nk}|<\tau} (x - \alpha_{nk})\, dF_{nk}(x)$$

$$= \sum_{k=1}^{k_n} \Bigl( \int_{|x-\alpha_{nk}|<\tau} x\, dF_{nk}(x) - \int_{|x|<\tau} x\, dF_{nk}(x) + \alpha_{nk} \int_{|x|\geqslant\tau} dF'_{nk}(x) \Bigr).$$

According to Theorem 2 of § 23 there exists a constant $C$ such that

$$\sum_{k=1}^{k_n} \int_{|x|\geqslant\tau} dF'_{nk}(x) \leqslant C. \,^*$$

From this, taking into account that

$$l_n = \sup_{1\leqslant k\leqslant k_n} |\alpha_{nk}| \to 0 \quad (n \to \infty), \tag{3}$$

we find

$$\Bigl| \sum_{k=1}^{k_n} \alpha_{nk} \int_{|x|\geqslant\tau} dF'_{nk}(x) \Bigr| \leqslant C l_n \to 0 \quad (n \to \infty).$$

On the other hand, for sufficiently large $n$,†

$$\Bigl| \sum_{k=1}^{k_n} \Bigl\{ \int_{|x-\alpha_{nk}|<\tau} x\, dF_{nk}(x) - \int_{|x|<\tau} x\, dF_{nk}(x) \Bigr|$$

$$\leqslant \sum_{k=1}^{k_n} \Bigl\{ \int_{|x+\tau|\leqslant|\alpha_{nk}|} x\, dF_{nk}(x) + \int_{|x-\tau|\leqslant|\alpha_{nk}|} x\, dF_{nk}(x) \Bigr\}$$

$$\leqslant 2\tau \sum_{k=1}^{k_n} \Bigl\{ \int_{|x+\tau|\leqslant 2|\alpha_{nk}|} dF'_{nk}(x) + \int_{|x-\tau|\leqslant 2|\alpha_{nk}|} dF'_{nk}(x) \Bigr\}.$$

---

* *Translator's note.* In order to prove the equivalence of 1′), 2′), 3′) and the conditions 1), 2) of the Theorem, we need the inequality $\sum_{k=1}^{k_n} \int_{|x|\geq\tau} dF'_{nk}(x) \leq C$ under each set of conditions. This inequality follows easily from either 1′) or 1), using the fact $\sup_{1\leq k\leq k_n} |\alpha_{nk}| \to 0$ $(n \to \infty)$ deduced at the beginning of the proof of Theorem 2 of § 23.

† *Translator's note.* The original formula after this sentence is incorrect and is corrected here in one of the possible ways.

The last member of this inequality approaches zero, by 1′), formula (3), and the fact that $-\tau$ and $+\tau$ are continuity points of the functions $M(u)$ and $N(u)$.

Thus (2) is proved.

Furthermore, at continuity points $u$ of the function $M(u)$

$$\underline{I}_n = \sum_{k=1}^{k_n} \int_{-\infty}^{u-l_n} dF_{nk}(x+\alpha_{nk}) \to M(u) \quad (n\to\infty),$$

$$\overline{I}_n = \sum_{k=1}^{k_n} \int_{-\infty}^{u+l_n} dF_{nk}(x+\alpha_{nk}) \to M(u) \quad (n\to\infty).$$

Now

$$\underline{I}_n \leqslant \sum_{k=1}^{k_n} \int_{-\infty}^{u} dF_{nk}(x) \leqslant \overline{I}_n,$$

and consequently,

$$\sum_{k=1}^{k_n} \int_{-\infty}^{u} dF_{nk}(x) = \sum_{k=1}^{k_n} F_{nk}(u) \to M(u) \quad (n\to\infty).$$

In exactly the same way it is proved that for $u > 0$

$$-\sum_{k=1}^{k_n} \int_{u}^{\infty} dF_{nk}(x) = \sum_{k=1}^{k_n} (F_{nk}(u)-1) \to N(u).$$

We have thereby proved that 1′) implies 1). The converse is proved just as simply, and we shall not pause for it.*

It remains to establish the equivalence of 2) and 2′).

For this purpose we note first of all the following obvious relation:

$$\lim_{\varepsilon\to 0} \overline{\lim_{n\to\infty}} \sum_{k=1}^{k_n} \int_{|x|<\varepsilon} x^2\, dF_{nk}(x+\alpha_{nk})$$

$$= \lim_{\varepsilon\to 0} \overline{\lim_{n\to\infty}} \sum_{k=1}^{k_n} \int_{|x+\alpha_{nk}|<\varepsilon} x^2\, dF_{nk}(x+\alpha_{nk})$$

$$= \lim_{\varepsilon\to 0} \overline{\lim_{n\to\infty}} \sum_{k=1}^{k_n} \int_{|x|<\varepsilon} (x-\alpha_{nk})^2\, dF_{nk}(x),$$

$$\lim_{\varepsilon\to 0} \underline{\lim_{n\to\infty}} \sum_{k=1}^{k_n} \int_{|x|<\varepsilon} x^2\, dF_{nk}(x+\alpha_{nk})$$

$$= \lim_{\varepsilon\to 0} \underline{\lim_{n\to\infty}} \sum_{k=1}^{k_n} \int_{|x|<\varepsilon} (x-\alpha_{nk})^2\, dF_{nk}(x).$$

---

* See the translator's note just before (3).

Furthermore,

$$\sum_{k=1}^{k_n}\int_{|x|<\varepsilon}(x-\alpha_{nk})^2\,dF_{nk}(x)=\sum_{k=1}^{k_n}\left\{\int_{|x|<\varepsilon}x^2\,dF_{nk}(x)\right.$$
$$\left.-2\alpha_{nk}\int_{|x|<\varepsilon}x\,dF_{nk}(x)+\alpha_{nk}^2-\alpha_{nk}^2\int_{|x|\geqslant\varepsilon}dF_{nk}(x)\right\}$$
$$=\sum_{k=1}^{k_n}\left\{\int_{|x|<\varepsilon}x^2\,dF_{nk}(x)-\left(\int_{|x|<\varepsilon}x\,dF_{nk}(x)\right)^2\right\}$$
$$+\sum_{k=1}^{k_n}\left\{\left(\int_{\varepsilon\leqslant|x|<\tau}x\,dF_{nk}(x)\right)^2-\alpha_{nk}^2\int_{|x|\geqslant\varepsilon}dF_{nk}(x)\right\}.$$

We denote the second sum of the last member of the equality above by $\omega_n$.

It is easy to see that

$$|\omega_n|\leqslant\left(\tau^2\cdot\sup_{1\leqslant k\leqslant k_n}\int_{|x|\geqslant\varepsilon}dF_{nk}(x)+l_n^2\right)\cdot\sum_{k=1}^{k_n}\int_{|x|\geqslant\varepsilon}dF_{nk}(x).$$

The first factor approaches zero as $n\to\infty$. The second factor is bounded.* In fact, if the distribution functions of the sums (1) of § 24 converge to a limit, then by Theorem 2 of § 23

$$\sum_{k=1}^{k_n}\int_{|x|>\varepsilon}dF_{nk}(x)\leqslant\sum_{k=1}^{k_n}\int_{|x|>\frac{\varepsilon}{2}}dF_{nk}(x+\alpha_{nk})$$
$$\leqslant\frac{4+\varepsilon^2}{\varepsilon^2}\sum_{k=1}^{k_n}\int_{|x|>\frac{\varepsilon}{2}}\frac{x^2}{1+x^2}\,dF(x+\alpha_{nk})\leqslant\frac{4+\varepsilon^2}{\varepsilon^2}C.$$

If, on the other hand, the conditions of Theorem 1 are satisfied, then, whatever $\delta>0$ is, for sufficiently large $n$

$$\sum_{k=1}^{k_n}\int_{|x|>\varepsilon}dF_{nk}(x)\leqslant M(-\varepsilon)-N(\varepsilon)+\delta.$$

Thus in both cases

$$\omega_n\to 0.$$

We have thereby completed the proof of the equivalence of the conditions 1′), 2′), 3′) and the conditions of the theorem. The theorem is proved completely.

---

* *Translator's note.* This follows also from 1′). See the preceding note.

*Remark.* Theorem 1 of § 19 and Theorem 1 of § 24 enable us to formulate the conditions of existence of a limit law for the sums (1) of § 24 in another way.

*In order that for some suitably chosen constants $A_n$ the distribution laws of the sums* (1) *of infinitesimal summands converge to a limit law, it is necessary and sufficient that there exist a nondecreasing function $G(u)$ of bounded variation such that*

$$\sum_{k=1}^{k_n} \int_{-\infty}^{u} \frac{x^2}{1+x^2}\, dF_{nk}(x+a_{nk}) \Rightarrow G(u)$$

*as $n \to \infty$.*

*The function $G(u)$ determines the limit distribution function according to the formula of Lévy and Khintchine.*

It is obvious that the results obtained above can be carried over automatically to asymptotically constant summands. We shall confine ourselves to formulating one theorem which is almost a literal repetition of Theorem 1.

THEOREM 2. *In order that for some suitably chosen constants $A_n$ the distribution laws of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n} - A_n$$

*of independent asymptotically constant random variables converge to a limit, it is necessary and sufficient that there exist functions $M(u)$ and $N(u)$ and a constant $\sigma$ such that:*

1) *At continuity points of the functions $M(u)$ and $N(u)$*

$$\sum_{k=1}^{k_n} \int_{-\infty}^{u} dF_{nk}(x+m_{nk}) \to M(u) \qquad (n \to \infty) \quad (u < 0),$$

$$\sum_{k=1}^{k_n} \int_{u}^{\infty} dF_{nk}(x+m_{nk}) \to -N(u) \quad (n \to \infty) \ (u > 0);$$

2) $$\lim_{\varepsilon \to 0} \varliminf_{n \to \infty} \sum_{k=1}^{k_n} \left\{ \int_{|x|<\varepsilon} x^2\, dF_{nk}(x+m_{nk}) - \left( \int_{|x|<\varepsilon} x\, dF_{nk}(x+m_{nk}) \right)^2 \right\}$$

$$= \lim_{\varepsilon \to 0} \varlimsup_{n \to \infty} \sum_{k=1}^{k_n} \left\{ \int_{|x|<\varepsilon} x^2\, dF_{nk}(x+m_{nk}) - \left( \int_{|x|<\varepsilon} x\, dF_{nk}(x+m_{nk}) \right)^2 \right\} = \sigma^2.$$

*The constants $A_n$ may be chosen according to the formula*

$$A_n = \sum_{k=1}^{k_n} \int_{|x|<\tau} x\, dF_{nk}(x+m_{nk}) + \sum_{k=1}^{k_n} m_{nk} - \gamma(\tau),$$

*where $\gamma(\tau)$ is any constant.*

*The logarithm of the characteristic function of the limit law is given by the formula* (7) of § 18.

This theorem is a consequence of the preceding one, since the variables $\xi_{nk} - m_{nk}$ are infinitesimal.

We consider the particular case of the last theorem when

$$\lim_{\varepsilon \to 0} \overline{\lim_{n\to\infty}} \sum_{k=1}^{k_n} \left( \int_{|x|<\varepsilon} x\, dF_{nk}(x+m_{nk})\right)^2 = 0.$$

This circumstance takes place whenever the limit law does not have a normal component ($\sigma = 0$) or whenever the variables $\xi_{nk}$ are symmetrical with respect to the medians.

THEOREM 3. If

$$\lim_{\varepsilon \to 0} \overline{\lim_{n\to\infty}} \sum_{k=1}^{k_n} \left( \int_{|x|<\varepsilon} x\, dF_{nk}(x+m_{nk})\right)^2 = 0, \tag{4}$$

*then for the convergence (for suitably chosen $A_n$) of the distribution laws of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \dots + \xi_{nk_n} - A_n \tag{1}$$

*of independent asymptotically constant random variables to a limit, it is necessary and sufficient that there exist a function $G(u)$ such that*

$$G_n(u) = \sum_{k=1}^{k_n} \int_{-\infty}^{u} \frac{x^2}{1+x^2}\, dF_{nk}(x+m_{nk}) \Rightarrow G(u)$$

*as $n \to \infty$. The constants $A_n$ may be chosen according to the formula*

$$A_n = \sum_{k=1}^{k_n} \left[ \int_{|x|<\tau} x\, dF_{nk}(x+m_{nk}) + m_{nk} \right].$$

*Proof.* If (4) is satisfied, then, by the preceding theorem, for the convergence of the distribution laws of the sums (1) to a limit it is necessary and sufficient that there exist a constant $\sigma$ and functions $M(u)$ and $N(u)$ such that as $n \to \infty$

1) $\displaystyle\sum_{k=1}^{k_n} F_{nk}(x+m_{nk}) \to M(x) \qquad (x<0),$

2) $\displaystyle\sum_{k=1}^{k_n} (F_{nk}(x+m_{nk})-1) \to N(x) \qquad (x>0),$

3) $$\lim_{\varepsilon\to 0} \overline{\lim_{n\to\infty}} \sum_{k=1}^{k_n} \int_{|x|<\varepsilon} x^2\, dF_{nk}(x+m_{nk})$$
$$= \lim_{\varepsilon\to 0} \underline{\lim_{n\to\infty}} \sum_{k=1}^{k_n} \int_{|x|<\varepsilon} x^2\, dF_{nk}(x+m_{nk}) = \sigma^2.$$

If we introduce the function $G(u)$, defined by the equations

$$G(u) = \int_{-\infty}^{u} \frac{x^2}{1+x^2}\, dM(x) \qquad (u < 0),$$

$$G(+0) = \sigma^2 + \int_{-\infty}^{0} \frac{x^2}{1+x^2}\, dM(x),$$

$$G(u) = G(+0) + \int_{0}^{u} \frac{x^2}{1+x^2}\, dN(x) \qquad (u > 0),$$

then by Theorems 1 and 2 of § 19 the relations 1)–3) just written down are equivalent to the condition $G_n \Rightarrow G$ in the theorem considered.

We shall now prove that if the limit law does not have a normal component (i.e., if $\sigma = 0$), then

$$\lim_{\epsilon \to 0} \overline{\lim_{n\to\infty}} \sum_{k=1}^{k_n} \Big( \int_{|x|<\epsilon} x\, dF_{nk}(x + m_{nk}) \Big)^2 = 0.$$

For this purpose we note that if we denote $A_k$ by that one of the intervals $(-\epsilon, 0)$ and $(0, \epsilon)$ for which the integral

$$\int_{A} x\, dF_{nk}(x + m_{nk})$$

has a greater absolute value, then

$$\sum_{k=1}^{k_n} \Big( \int_{|x|<\epsilon} x\, dF_{nk}(x+m_{nk}) \Big)^2 \leqslant \sum_{k=1}^{k_n} \Big( \int_{A_k} x\, dF_{nk}(x+m_{nk}) \Big)^2$$
$$\leqslant \sum_{k=1}^{k_n} \int_{A_k} x^2\, dF_{nk}(x+m_{nk}) \int_{A_k} dF_{nk}(x+m_{nk})$$
$$\leqslant \sum_{k=1}^{k_n} \int_{A_k} dF_{nk}(x+m_{nk}) \int_{|x|<\epsilon} x^2\, dF_{nk}(x+m_{nk}).$$

But by the definition of the median

$$\int_{A_k} dF_{nk}(x+m_{nk}) \leqslant \frac{1}{2},$$

so that

$$\sum_{k=1}^{k_n} \Big( \int_{|x|<\epsilon} x\, dF_{nk}(x+m_{nk}) \Big)^2 \leqslant \frac{1}{2} \sum_{k=1}^{k_n} \int_{|x|<\epsilon} x^2\, dF_{nk}(x+m_{nk}).$$

Now the desired relation follows from Theorem 2 and from the inequality

$$\frac{1}{2}\sum_{k=1}^{k_n}\int_{|x|<\varepsilon} x^2 dF_{nk}(x+m_{nk}) \leqslant \sum_{k=1}^{k_n}\left\{\int_{|x|<\varepsilon} x^2 dF_{nk}(x+m_{nk}) - \left(\int_{|x|<\varepsilon} x\, dF_{nk}(x+m_{nk})\right)^2\right\}.$$

Modifying somewhat the formulation of the last three theorems we can obtain not only conditions for the existence of a limit law for the sums, but also conditions for the convergence to any given limit law. Let us paraphrase, for example, Theorem 1.

THEOREM 4. *In order that for suitably chosen constants $A_n$ the distribution functions of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \dots + \xi_{nk_n} - A_n$$

*of independent infinitesimal random variables converge to the distribution function $F(x)$, it is necessary and sufficient that the following conditions be satisfied:*

1) *At continuity points of $M(u)$ and $N(u)$*

$$\sum_{k=1}^{k_n} F_{nk}(x) \to M(x) \qquad \text{for } x < 0,$$

$$\sum_{k=1}^{k_n} (F_{nk}(x) - 1) \to N(x) \qquad \text{for } x > 0$$

*as $n \to \infty$;*

2) $$\lim_{\varepsilon\to 0}\varlimsup_{n\to\infty}\sum_{k=1}^{k_n}\left\{\int_{|x|<\varepsilon} x^2 dF_{nk}(x) - \left(\int_{|x|<\varepsilon} x\, dF_{nk}(x)\right)^2\right\}$$

$$= \lim_{\varepsilon\to 0}\varliminf_{n\to\infty}\sum_{k=1}^{k_n}\left\{\int_{|x|<\varepsilon} x^2 dF_{nk}(x) - \left(\int_{|x|<\varepsilon} x\, dF_{nk}(x)\right)^2\right\} = \sigma^2,$$

*where the functions $M(u)$, $N(u)$ and the constant $\sigma^2$ are determined by Lévy's formula for $F(x)$. The constants $A_n$ are determined by*

$$A_n = \sum_{k=1}^{k_n}\int_{|x|<\tau} x\, dF_{nk}(x) - \gamma_n(\tau),$$

*where $\gamma_n(\tau)$ is any convergent sequence of real numbers.*

*Remark.* If it is required to state the condition for the convergence of the distribution functions of the sums

$$\zeta_n = \xi_{n1} + \xi_{n2} + \dots + \xi_{nk_n} \tag{5}$$

to a limit distribution function, then one more condition should be added to the conditions of Theorem 4:

$$\lim_{n\to\infty}\sum_{k=1}^{k_n}\int_{|x|<\tau} x\, dF_{nk}(x) = \gamma(\tau).$$

# CHAPTER 5

# CONVERGENCE TO NORMAL, POISSON, AND UNITARY DISTRIBUTIONS

## § 26. Conditions for Convergence to Normal and Poisson Laws

We shall now make use of the general theorems in the preceding chapter to clarify conditions for convergence of distribution functions of sums to the various particular limit laws with which the classical theory of probability concerned itself. In this connection we shall confine ourselves in almost all theorems to the consideration of sums of infinitesimal random variables, since the consideration of asymptotically constant summands, as we have seen before, can be reduced to that of infinitesimal ones. And only in the treatment of theorems of the type of the law of large numbers does the very essence of the problem compel us to consider asymptotically constant summands.

The general problems considered in the preceding chapter were raised and solved only in recent years. The main interest of classical investigations amounted to the clarification of conditions for the convergence of distribution functions of sums to the normal law and to the determination of the broadest conditions under which the law of large numbers holds. It is interesting to note that, in essence, the classical theory of probability studied only one proper limit distribution law — the normal law. The study of the Poisson law was confined only to elementary investigations. The causes for such one-sidedness of classical investigations have been completely uncovered in recent times. It turns out that the normal distribution law indeed plays a dominating role in theoretical as well as applied questions. We shall see below that whereas for the convergence of distribution functions of sums of independent variables to the normal law only restrictions of a very general kind, apart from that of being infinitesimal (or asymptotically constant), have to be imposed on the summands, for the convergence to another limit law some very special properties are required of the summands.

We see that the general theorems developed in the preceding chapters permit us to obtain, literally in a few words, the proofs of the most impor-

tant theorems in the theory of probability. However, this circumstance must not belittle in the eyes of the reader either the value of those theorems themselves or that of the efforts spent by mathematicians in the formulation and proof of those propositions.

We now turn to the discussion of concrete results, and we begin this discussion with the proof of a theorem of A. Ya. Khintchine [59], clarifying the fundamental importance of the normal law in the theory of probability.

THEOREM 1. *If the distributions of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n}$$

*of infinitesimal random variables $\xi_{nk}$ $(1 \leq k \leq k_n)$ which are independent in each row converge to a limit, then the relation*

$$\sum_{k=1}^{k_n} \int_{|x| \geqslant \epsilon} dF_{nk}(x) \to 0 \qquad (n \to \infty) \tag{1}$$

*is satisfied for every $\epsilon > 0$ if and only if the limit law is normal.*

*Proof.* Since by hypothesis a limit law exists, we have, by Theorem 4 of § 25,

$$1) \quad \sum_{k=1}^{k_n} F_{nk}(x) = \sum_{k=1}^{k_n} \int_{-\infty}^{x} dF_{nk}(z) \to M(x) \qquad (x < 0),$$

$$2) \quad \sum_{k=1}^{k_n} (F_{nk}(x) - 1) = -\sum_{k=1}^{k_n} \int_{x}^{\infty} dF_{nk}(z) \to N(x) \qquad (x > 0).$$

Hence we conclude that if (1) is satisfied, then $M(x) \equiv 0$, $N(x) \equiv 0$, and consequently the limit law is normal. Conversely, if it is known that the limit law is normal, then $M(x) \equiv 0$, $N(x) \equiv 0$, and therefore by 1) and 2) the relation (1) holds, proving the theorem.

We have seen that if we impose on the variables $\xi_{nk}$ only the requirement of being infinitesimal, i.e., the requirement that as $n \to \infty$

$$\sup_{1 \leqslant k \leqslant k_n} \mathsf{P}\{|\xi_{nk}| \geqslant \varepsilon\} \to 0 \tag{2}$$

for every $\epsilon > 0$, then any infinitely divisible law can serve as the limit law for the sums $\zeta_n$. The condition (1) means, as we shall now prove, not only that the individual summands are small, but that they are uniformly small. In other words, for every $\epsilon > 0$ the probability that at least one of the $\xi_{nk}$ $(1 \leq k \leq k_n)$ exceeds $\epsilon$ approaches zero as $n \to \infty$. To put it in a formula, this can be written as

$$\mathsf{P}\{\sup_{1\leqslant k\leqslant k_n}|\xi_{nk}|\geqslant\varepsilon\}\to 0 \,^{*} \qquad (n\to\infty). \tag{3}$$

It is clear that

$$\mathsf{P}\{\sup_{1\leqslant k\leqslant k_n}|\xi_{nk}|\geqslant\varepsilon\}=1-\mathsf{P}\{\sup_{1\leqslant k\leqslant k_n}|\xi_{nk}|<\varepsilon\}$$

$$=1-\prod_{k=1}^{k_n}\mathsf{P}\{|\xi_{nk}|<\varepsilon\}=1-\prod_{k=1}^{k_n}\Big(1-\int_{|x|\geqslant\varepsilon}dF_{nk}(x)\Big).$$

Hence the condition (3) and

$$\prod_{k=1}^{k_n}\Big(1-\int_{|x|\geqslant\varepsilon}dF_{nk}(x)\Big)\to 1 \qquad (n\to\infty) \tag{4}$$

are equivalent.

The inequality

$$1-\sum_{k=1}^{k_n}\int_{|x|\geqslant\varepsilon}dF_{nk}(x)\leqslant\prod_{k=1}^{k_n}\Big(1-\int_{|x|\geqslant\varepsilon}dF_{nk}(x)\Big)\leqslant e^{-\sum_{k=1}^{k_n}\int_{|x|\geqslant\varepsilon}dF_{nk}(x)}\leqslant 1$$

shows that (1) implies (4) and so also (3), and conversely that (3) implies (1).

The preceding theorem has its analogue in the theory of stochastic processes with independent increments. The collection of random variables $\zeta_\lambda$ depending on the continuously varying real parameter $\lambda$ is called a stochastic process with independent increments, if the increments of the random variable $\zeta_\lambda$ in disjoint intervals of the parameter $\lambda$ are independent random variables.

---

* We remark that the random variables can be infinitesimal without satisfying the requirement (3).

For example, let

$$F_{nk}(x)=\begin{cases}0 & \text{for } x\leqslant 0,\\ 1-\dfrac{1}{n} & \text{for } 0<x\leqslant 1\\ 1 & \text{for } x>1.\end{cases} \qquad (1\leqslant k\leqslant n),$$

Then for every $\epsilon$ $(0<\epsilon<1)$, as $n\to\infty$

$$\sup_{1\leqslant k\leqslant k_n}\mathsf{P}\{|\xi_{nk}|>\varepsilon\}=\frac{1}{n}\to 0,$$

but at the same time

$$\mathsf{P}\{\sup_{1\leqslant k\leqslant k_n}|\xi_{nk}|>\varepsilon\}=1-\prod_{k=1}^{k_n}\mathsf{P}\{|\xi_{nk}|<\varepsilon\}$$

$$=1-\Big(1-\frac{1}{n}\Big)^n\to 1-\frac{1}{e}\ (n\to\infty).$$

We shall say that the stochastic process $\zeta_\lambda$ is stochastically continuous, if for every $\epsilon > 0$

$$\mathsf{P}\{|\zeta_{\lambda+\Delta\lambda} - \zeta_\lambda| > \varepsilon\} \to 0 \qquad (\Delta\lambda \to 0).$$

Furthermore, we shall say that the process $\zeta_\lambda$ is stochastically *strongly* continuous in the interval $[\lambda_0, \Lambda]$, if for every sequence $\lambda_0 \leq \lambda_1 < \lambda_2 < \cdots < \lambda_n = \Lambda$ and arbitrary $\epsilon > 0$

$$\mathsf{P}\{\sup_{1 \leqslant k \leqslant n} |\zeta_{\lambda_k} - \zeta_{\lambda_{k-1}}| > \varepsilon\} \to 0$$

as $\max(\lambda_k - \lambda_{k-1}) \to 0$.

Now Theorem 1 can be formulated in the language of stochastic processes as follows:

*In order that the increments $\zeta_{\lambda_2} - \zeta_{\lambda_1}$ of a stochastic process with independent increments in the interval $\lambda_0 \leq \lambda_1 < \lambda_2 \leq \Lambda$ be normally distributed, it is necessary and sufficient that the process $\zeta_\lambda$ be stochastically strongly continuous in the interval* $[\lambda_0, \Lambda]$.

From the point of view of the developed theory of stochastic processes, strong continuity of $\zeta_\lambda$ means that *with probability one $\zeta_\lambda$ as a function of* $\lambda$ *is continuous at all points* $\lambda$. However, we cannot go into the foundation of this assertion here (cf. § 16 and Ch. VIII of the book [76]).

alm. all samp fns contn ?

THEOREM 2. *In order that for some suitably chosen constants $A_n$ the distributions of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \dots + \xi_{nk_n} - A_n$$

*converge as $n \to \infty$ to the normal law*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz \tag{5}$$

*and the summands $\xi_{nk}$ $(1 \leq k \leq k_n)$ be infinitesimal, it is necessary and sufficient that the conditions*

$$1) \quad \sum_{k=1}^{k_n} \int_{|x| \geqslant \epsilon} dF_{nk}(x) \to 0,$$

$$2) \quad \sum_{k=1}^{k_n} \left\{ \int_{|x| < \varepsilon} x^2 dF_{nk}(x) - \left( \int_{|x| < \varepsilon} x \, dF_{nk}(x) \right)^2 \right\} \to 1$$

*be satisfied for every $\epsilon > 0$, as $n \to \infty$.*

*Proof. Sufficiency.* The condition 1) of the theorem implies that the variables $\xi_{nk}$ are infinitesimal. In fact, for every $\epsilon > 0$ we have, as $n \to \infty$,

$$\sup_{1 \leqslant k \leqslant k_n} \mathsf{P}\{|\xi_{nk}| > \varepsilon\} = \sup_{1 \leqslant k \leqslant k_n} \int_{|x| > \epsilon} dF_{nk}(x)$$
$$\leqslant \sum_{k=1}^{k_n} \int_{|x| > \varepsilon} dF_{nk}(x) \to 0.$$

Therefore we find ourselves under the conditions of the preceding section and can make use of the general results obtained there.

The necessity of the conditions of the theorem is obtained from Theorem 4 of § 25, if we put there $a = 1$, $\gamma(\tau) = 0$, and $M(-u) \equiv N(u) \equiv 0$.

In fact, the first condition of Theorem 4 of § 25 can be written as follows: For every $x > 0$

$$\sum_{k=1}^{k_n} \int_{|u|>x} dF_{nk}(u) \leqslant^{\dagger} \sum_{k=1}^{k_n} (1 - F_{nk}(x) + F_{nk}(-x)) \\ \to M(-x) + N(x) = 0.$$

It is clear that this condition coincides with 1).

Furthermore, for every $\epsilon'$ $(0 < \epsilon' < \epsilon)$, we have

$$\begin{aligned} \sum_{k=1}^{k_n} & \left\{ \int_{|x|<\epsilon} x^2\, dF_{nk}(x) - \left( \int_{|x|<\epsilon} x\, dF_{nk}(x) \right)^2 \right\} \\ &= \sum_{k=1}^{k_n} \left\{ \int_{|x|<\epsilon'} x^2\, dF_{nk}(x) - \left( \int_{|x|<\epsilon'} x\, dF_{nk}(x) \right)^2 \right\} \\ &+ \sum_{k=1}^{k_n} \left\{ \int_{\epsilon' \leqslant |x| < \epsilon} x^2\, dF_{nk}(x) - \left( \int_{\epsilon' \leqslant |x| < \epsilon} x\, dF_{nk}(x) \right)^2 \right\} \\ &\qquad - 2 \sum_{k=1}^{k_n} \left( \int_{|x|<\epsilon'} x\, dF_{nk}(x) \right) \left( \int_{\epsilon' \leqslant |x| \leqslant \epsilon} x\, dF_{nk}(x) \right). \end{aligned}$$

Now

$$\begin{aligned} 0 &\leqslant \sum_{k=1}^{k_n} \left\{ \int_{\epsilon' \leqslant |x| < \epsilon} x^2\, dF_{nk}(x) - \left( \int_{\epsilon' \leqslant |x| < \epsilon} x\, dF_{nk}(x) \right)^2 \right\} \\ &\leqslant \sum_{k=1}^{k_n} \int_{\epsilon' \leqslant |x| < \epsilon} x^2\, dF_{nk}(x) \leqslant \epsilon^2 \sum_{k=1}^{k_n} \int_{\epsilon' \leqslant |x| < \epsilon} dF_{nk}(x) \\ &\qquad \leqslant \epsilon^2 \sum_{k=1}^{k_n} \int_{|x|>\epsilon'} dF_{nk}(x) \end{aligned}$$

and

$$2 \sum_{k=1}^{k_n} \left| \int_{|x|<\epsilon'} x\, dF_{nk}(x) \right| \left| \int_{\epsilon' \leqslant |x| < \epsilon} x\, dF_{nk}(x) \right| \\ \leqslant 2\epsilon' \cdot \epsilon \sum_{k=1}^{k_n} \int_{|x| \geqslant \epsilon'} dF_{nk}(x).$$

This last sum tends to zero as $n \to \infty$, by the first condition of the theorem already proved. Thus, for every $\epsilon > 0$ and $\epsilon' > 0$,

---

† *Translator's note.* In the original, an equality sign stands here. This is incorrect unless $x$ is a continuity point for all $F_{nk}(x)$.

$$\varlimsup_{n\to\infty} \sum_{k=1}^{k_n} \left\{ \int\limits_{|x|<\epsilon} x^2\, dF_{nk}(x) - \Big( \int\limits_{|x|<\epsilon} x\, dF_{nk}(x) \Big)^2 \right\}$$
$$= \varlimsup_{n\to\infty} \sum_{k=1}^{k_n} \left\{ \int\limits_{|x|<\epsilon'} x^2\, dF_{nk}(x) - \Big( \int\limits_{|x|<\epsilon'} x\, dF_{nk}(x) \Big)^2 \right\},$$

i.e., the upper limit does not depend on $\epsilon$. This is also true of the lower limit. Hence by the conditions of Theorem 4 of § 25 we conclude that for every $\epsilon > 0$ not only the upper and lower limits of the expression

$$\sum_{k=1}^{k_n} \left\{ \int\limits_{|x|<\epsilon} x^2\, dF_{nk}(x) - \Big( \int\limits_{|x|<\epsilon} x\, dF_{nk}(x) \Big)^2 \right\}$$

exist, but that also the ordinary limit exists.

For later purposes it will be important to write the conditions for convergence to the normal law in some other forms.

THEOREM 3. *In order that for some suitably chosen constants $A_n$ the distributions of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \dots + \xi_{nk_n} - A_n$$

*of independent infinitesimal random variables converge to the normal law (5), it is necessary and sufficient that for every $\epsilon > 0$*

$$1) \quad \sum_{k=1}^{k_n} \int\limits_{|x|>\epsilon} dF_{nk}(x+\alpha_{nk}) \to 0 \qquad (n\to\infty),$$

$$2) \quad \sum_{k=1}^{k_n} \int\limits_{|x|<\epsilon} x^2\, dF_{nk}(x+\alpha_{nk}) \to 1 \qquad (n\to\infty),$$

*where*
$$\alpha_{nk} = \int\limits_{|x|<\tau} x\, dF_{nk}(x),$$
*and $\tau$ is any positive number.*

We shall not give the proof of this theorem, since it is deduced from the conditions 1′) and 2′) of § 25 in the same way as the preceding theorem from Theorem 4 of § 25.

We shall present one more theorem concerning the convergence of distribution laws of sums of independent variables to the normal law.

The sufficiency of its conditions was indicated as early as in 1926 by S. N. Bernstein [5]. The complete theorem was proved by Feller [27] in 1935.

THEOREM 4. *In order that for a given sequence of independent random variables*

$$\xi_1, \xi_2, \dots, \xi_n, \dots$$

*it should be possible to find real constants $A_n$ and $B_n > 0$ having the property that the distribution laws of the sums*

$$\zeta_n = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{B_n} - A_n \tag{6}$$

*converge to the normal law* (5) *and the summands*

$$\xi_{nk} = \frac{\xi_k}{B_n}, \quad 1 \leqslant k \leqslant k_n,$$

*be infinitesimal, it is necessary and sufficient that there exist a sequence of constants $C_n (C_n \to \infty)$ such that as $n \to \infty$*

$$\left.\begin{aligned} &\sum_{k=1}^{n} \int_{|x| > C_n} dF_k(x) \to 0, \\ &\frac{1}{C_n^2} \sum_{k=1}^{n} \left\{ \int_{|x| < C_n} x^2 \, dF_k(x) - \left( \int_{|x| < C_n} x \, dF_k(x) \right)^2 \right\} \to \infty . \end{aligned}\right\} \tag{7}$$

*Proof. Necessity.* Since by hypothesis the random variables $\xi_{nk} = \dfrac{\xi_k}{B_n}$ are infinitesimal, we can make use of Theorem 2. Since $F_{nk}(x) = F_k(B_n x)$, the conditions 1) and 2) now take the form

1) $\displaystyle\sum_{k=1}^{n} \int_{|x| > \epsilon B_n} dF_k(x) \to 0,$

2) $\displaystyle\frac{1}{B_n^2} \sum_{k=1}^{n} \left\{ \int_{|x| < \epsilon B_n} x^2 \, dF_k(x) - \left( \int_{|x| < \epsilon B_n} x \, dF_k(x) \right)^2 \right\} \to 1 \ (n \to \infty).$

Obviously, we can pick a sequence

$$\varepsilon_n \to 0,$$

so that $\epsilon_n B_n \to \infty$ and

$$\left.\begin{aligned} &\sum_{k=1}^{n} \int_{|x| > \epsilon_n B_n} dF_k(x) \to 0, \\ &\frac{1}{B_n^2} \sum_{k=1}^{n} \left\{ \int_{|x| < \epsilon_n B_n} x^2 \, dF_k(x) - \left( \int_{|x| < \epsilon_n B_n} x \, dF_n(x) \right)^2 \right\} \to 1 \end{aligned}\right\} \tag{8}$$

as $n \to \infty$. Putting here $\epsilon_n B_n = C_n$, we obtain (7).

*Sufficiency.* Now let (7) be satisfied.

Put

$$B_n^2 = \sum_{k=1}^{n} \left\{ \int_{|x| < C_n} x^2 \, dF_k(x) - \left( \int_{|x| < C_n} x \, dF_k(x) \right)^2 \right\}.$$

Comparing this equation with the second condition of the theorem, we conclude that

$$C_n = o(B_n).$$

Hence for every $\epsilon > 0$ and sufficiently large $n$

$$\sum_{k=1}^{n} \int_{|x| > C_n} dF_k(x) \geqslant \sum_{k=1}^{n} \int_{|x| > \epsilon B_n} dF_k(x).$$

Thus, by the first condition of the theorem, for every $\epsilon > 0$

$$\sum_{k=1}^{n} \int_{|x| > \epsilon B_n} dF_k(x) \to 0 \tag{9}$$

as $n \to \infty$. Furthermore, we have

$$\frac{1}{B_n^2} \sum_{k=1}^{n} \left\{ \int_{C_n \leqslant |x| \leqslant \epsilon B_n} x^2\, dF_k(x) - \left( \int_{C_n \leqslant |x| < \epsilon B_n} x\, dF_k(x) \right)^2 \right\}$$

$$\leqslant \epsilon^2 \sum_{k=1}^{n} \int_{|x| \geqslant C_n} dF_k(x) \to 0, \quad n \to \infty$$

and

$$\frac{1}{B_n^2} \sum_{k=1}^{n} \left| \int_{|x| < C_n} x\, dF_k(x) \right| \left| \int_{C_n \leqslant |x| < \epsilon B_n} x\, dF_k(x) \right|$$

$$\leqslant \frac{\epsilon B_n \cdot C_n}{B_n^2} \sum_{k=1}^{n} \int_{|x| \geqslant C_n} dF_k(x) \to 0.$$

From this and from the definition of $B_n$, we see that for every $\epsilon > 0$

$$\frac{1}{B_n^2} \sum_{k=1}^{n} \left\{ \int_{|x| < \epsilon B_n} x^2\, dF_k(x) - \left( \int_{|x| < \epsilon B_n} x\, dF_k(x) \right)^2 \right\} \to 1 \tag{10}$$

as $n \to \infty$. Since (9) implies that the variables $\xi_k/B_n$, $(1 \leq k \leq n)$ are infinitesimal, we find ourselves under the conditions of Theorem 2; hence it follows that (9) and (10), and so also the conditions of the theorem, are sufficient for the convergence of the distribution laws of the sums to a normal law.

THEOREM 5.* *In order that the distribution laws of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n}$$

*of independent infinitesimal random variables converge to the Poisson law*

$$P(x) = \sum_{0 \leqslant k < x} \frac{e^{-\lambda} \lambda^k}{k!} \qquad (\lambda > 0),$$

---

* B. V. Gnedenko [36], J. Marcinkiewicz [81].

*it is necessary and sufficient that for every $\epsilon$ $(0 < \epsilon < 1)$ the following conditions be satisfied:*

$$1)\ \sum_{k=1}^{k_n} \int_{R_\epsilon} dF_{nk}(x) \to 0 \qquad (n \to \infty),$$

$$2)\ \sum_{k=1}^{k_n} \int_{|x-1|<\epsilon} dF_{nk}(x) \to \lambda \qquad (n \to \infty),$$

$$3)\ \sum_{k=1}^{k_n} \int_{|x|<\epsilon} x\, dF_{nk}(x) \to 0 \qquad (n \to \infty),$$

$$4)\ \sum_{k=1}^{k_n} \left\{ \int_{|x|<\epsilon} x^2 dF_{nk}(x) - \left( \int_{|x|<\epsilon} x\, dF_{nk}(x) \right)^2 \right\} \to 0 \qquad (n \to \infty),$$

*where $R_\epsilon$ denotes the domain obtained from the real line $-\infty < x < +\infty$ by discarding the intervals $|x| < \epsilon$ and $|x-1| < \epsilon$.*

*Proof.* This theorem is as easily deduced from Theorem 4 of § 25 as is Theorem 2; to this end it is sufficient to note that in Lévy's formula for the Poisson law we should put $M(u) \equiv 0$, $N(u) = -\lambda$ for $0 < u \leq 1$; $N(u) = 0$ for $u > 1$; $\sigma = 0$, $\gamma(\tau) = 0$ for $0 < \tau < 1$.

## § 27. The Law of Large Numbers

In § 22 the following theorem was obtained [see (10)].

Theorem 1. *In order that the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n} - A_n \tag{1}$$

*of random variables which are independent in each row obey the law of large numbers, the condition*

$$\sum_{k=1}^{k_n} \int \frac{x^2}{1+x^2} dF_{nk}(x + m_{nk}) \to 0 \qquad (n \to \infty).$$

*is necessary and sufficient. The constants $A_n$ may be chosen according to the formula*

$$A_n = \sum_{k=1}^{k_n} \left\{ m_{nk} + \int_{|x|<\tau} x\, dF_{nk}(x + m_{nk}) \right\},$$

*where $\tau$ is a constant.*

In the remark after Theorem 1 of § 23 it was indicated that we could obtain the results concerning the law of large numbers from the general theorems.

Now we are in a position to do so. Namely, Theorem 1 follows easily from Theorem 3 of § 25, if we take into account that for the unitary law we should put

$$\gamma = 0, \quad G(u) \equiv 0$$

in the formula of Lévy and Khintchine.

*Remark* 1. It is possible to give a somewhat different formulation of Theorem 1: *In order that the sums* (1) *obey the law of large numbers, it is necessary and sufficient that as* $n \to \infty$

$$1) \qquad \sum_{k=1}^{k_n} \int_{|x| \geqslant 1} dF_{nk}(x + m_{nk}) \to 0,$$

$$2) \qquad \sum_{k=1}^{k_n} \int_{|x| < 1} x^2 \, dF_{nk}(x + m_{nk}) \to 0.$$

*Moreover, we may take*

$$A_n = \sum_{k=1}^{k_n} \left\{ \int_{|x| < 1} x \, dF_{nk}(x + m_{nk}) + m_{nk} \right\}.$$

*Remark* 2. When it is also required that $A_n = 0$, and consequently the question concerns the conditions under which

$$\mathsf{P}\{|\xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n}| > \varepsilon\} \to 0 \qquad (n \to \infty)$$

for every $\epsilon > 0$, then the conditions given above are not yet sufficient, and it is necessary to add the new requirement that for every $\tau > 0$

$$\sum_{k=1}^{k_n} \left\{ \int_{|x| < \tau} x \, dF_{nk}(x + m_{nk}) + m_{nk} \right\} \to 0 \qquad (n \to \infty).$$

*Remark* 3. *From Theorem* 1* *of* § 25 *we deduce: In order that the*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n}$$

*of independent random variables converge in probability to zero as* $n \to \infty$ *and the variables* $\xi_{nk}$ $(1 \leq k \leq k_n)$ *be infinitesimal, it is necessary and sufficient that for every* $\epsilon > 0$ *the following relations be satisfied as* $n \to \infty$:

$$1) \qquad \sum_{k=1}^{k_n} \int_{|x| > \varepsilon} dF_{nk}(x) \to 0,$$

$$2) \qquad \sum_{k=1}^{k_n} \int_{|x| < \varepsilon} x \, dF_{nk}(x) \to 0,$$

$$3) \qquad \sum_{k=1}^{k_n} \left\{ \int_{|x| < \varepsilon} x^2 \, dF_{nk}(x) - \left( \int_{|x| < \varepsilon} x \, dF_{nk}(x) \right)^2 \right\} \to 0.$$

---

* *Translator's note.* Rather, Theorem 4 of § 25 and the Remark after it, noting that $\dot{\gamma}(\tau) = 0$.

As a simple particular case of Theorem 1, we state the following theorem, proved under the hypothesis $B_n = n$ by A. N. Kolmogorov [63] and in the general form by W. Feller [28].

**Theorem 2.** *In order that the sequence*

$$\xi_1, \quad \xi_2, \ldots, \quad \xi_n, \ldots$$

*of independent random variables obey the law of large numbers*, i.e., *that for a given sequence $B_n > 0$ there exist constants $A_n$ such that for every $\epsilon > 0$*

$$\mathsf{P}\left\{\left|\frac{\xi_1+\xi_2+\ldots+\xi_n}{B_n} - A_n\right| > \varepsilon\right\} \to 0 \qquad (n\to\infty),$$

*it is necessary and sufficient that*

$$\sum_{k=1}^{n}\int \frac{x^2}{B_n^2+x^2}\, dF_k(x+m_k) \to 0 \qquad (n\to\infty),$$

*where $m_k$ is a median of the variable $\xi_k$.*

*The constants $A_n$ may be chosen according to the formula*

$$A_n = \frac{1}{B_n}\sum_{k=1}^{n}\Bigl(m_k + \int\limits_{|x|<\tau B_n} x\, dF_k(x+m_k)\Bigr), \tag{2}$$

*where $\tau$ is an arbitrary positive number.*

In addition to the theorems given above, we shall prove the theorem establishing necessary and sufficient conditions for the validity of the law of large numbers in its classical formulation.

**Theorem 3.*** *In order that the sequence of independent random variables $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ having finite mathematical expectations $\mathbf{M}\xi_k = a_k$ obey the law of large numbers*, i.e., *for every $\epsilon > 0$*

$$\mathsf{P}\left\{\left|\frac{\sum_{k=1}^{n}(\xi_k - a_k)}{n}\right| \geqslant \varepsilon\right\} \to 0 \qquad (n\to\infty), \tag{3}$$

*it is necessary and sufficient that as $n\to\infty$*

1) $$\sum_{k=1}^{n}\int\limits_{|x|>n} dF_k(x+a_k) \to 0,$$

2) $$\sum_{k=1}^{n}\frac{1}{n}\int\limits_{|x|<n} x\, dF_k(x+a_k) \to 0,$$

3) $$\sum_{k=1}^{n}\frac{1}{n^2}\int\limits_{|x|<n} x^2\, dF_k(x+a_k) \to 0.$$

---

* See [63].

*Proof.* For simplicity, we put

$$\xi_k' = \xi_k - a_k$$

We shall prove that if (3) is satisfied, i.e., if for every $\epsilon > 0$

$$\mathsf{P}\left\{\left|\frac{\xi_1' + \xi_2' + \ldots + \xi_n'}{n}\right| \geqslant \varepsilon\right\} \to 0 \quad (n \to \infty), \tag{4}$$

then the variables $\frac{\xi_k'}{n}$ $(1 \leq k \leq n)$ are infinitesimal.

We have for every $\delta > 0$ and for sufficiently large $n$

$$\mathsf{P}\left\{\left|\frac{\xi_1' + \ldots + \xi_{n-1}'}{n-1}\right| \leqslant \varepsilon\right\} > 1 - \delta \tag{5}$$

and

$$\mathsf{P}\left\{\left|\frac{\xi_1' + \ldots + \xi_n'}{n}\right| \leqslant \varepsilon\right\} > 1 - \delta. \tag{6}$$

Consequently,

$$\mathsf{P}\left\{\left|\frac{\xi_n}{n}\right| \leqslant 2\varepsilon\right\}$$
$$\geqslant \mathsf{P}\left\{\left(\left|\frac{\xi_1 + \ldots + \xi_{n-1}}{n-1}\right| \leqslant \varepsilon\right) \cap \left(\left|\frac{\xi_1 + \ldots + \xi_n}{n}\right| \leqslant \varepsilon\right)\right\} \geqslant 1 - 2\delta,$$

which is equivalent to saying that the variables

$$\frac{\xi_k}{n},\ 1 \leqslant k \leqslant n$$

are infinitesimal.

We can therefore make use of the assertion formulated in Remark 3 to Theorem 1. Then we obtain that for (4) it is necessary and sufficient that for every $\epsilon > 0$, as $n \to \infty$,

$$\sum_{k=1}^{n} \int_{|x| > \epsilon} dF_k'(nx) \to 0, \tag{7}$$

$$\sum_{k=1}^{n} \int_{|x| < \epsilon} x\, dF_k'(nx) \to 0, \tag{8}$$

$$\sum_{k=1}^{n} \left\{\int_{|x| < \varepsilon} x^2\, dF_k'(nx) - \left(\int_{|x| < \epsilon} x\, dF_k'(nx)\right)^2\right\} \to 0. \tag{9}$$

Since the mathematical expectations exist, as $n \to \infty$

$$\int_{|x| < \epsilon n} x\, dF_k'(x) \to \int x\, dF_k'(x) = 0.$$

Hence we conclude that

$$\sum_{k=1}^{n}\Big(\int_{|x|<\epsilon} x\,dF_k'(nx)\Big)^2 = \frac{1}{n^2}\sum_{k=1}^{n}\Big(\int_{|x|<\epsilon n} x\,dF_k'(x)\Big)^2 \to 0 \qquad (n\to\infty)$$

and that consequently (9) may be replaced by

$$\sum_{k=1}^{n}\int_{|x|<\epsilon} x^2\,dF_k'(nx) \to 0 \quad (n\to\infty). \tag{10}$$

Now the equivalence of the first and third conditions of the theorem with (7) and (9) is obvious. It remains to prove that

$$\sum_{k=1}^{n}\int_{|x|<\epsilon} x\,dF_k'(nx) - \sum_{k=1}^{n}\frac{1}{n}\int_{|x|<n} x\,dF_k'(x) \to 0 \qquad (n\to\infty).$$

But this equation follows from (7); in fact,

$$\Big|\sum_{k=1}^{n}\Big\{\int_{|x|<1} x\,dF_k'(nx) - \int_{|x|<\epsilon} x\,dF_k'(nx)\Big\}\Big|$$

$$= \Big|\sum_{k=1}^{n}\int_{\epsilon\leqslant|x|<1} x\,dF_k'(nx)\Big| \leqslant \sum_{k=1}^{n}\int_{|x|\geqslant\epsilon} dF_k'(nx) \to 0 \quad (n\to\infty).$$

The preceding results permit us to obtain the following interesting corollary.

Corollary 1. *In order that the series*

$$\sum_{k=1}^{\infty}\xi_k \tag{11}$$

*of independent random variables converge with probability one, it is necessary and sufficient that the "Cauchy criterion" be satisfied: for every* $\epsilon>0$ *and for* $m>n$

$$\mathsf{P}\{|\xi_n+\xi_{n+1}+\dots+\xi_m|>\epsilon\}\to 0 \qquad (n\to\infty).$$

*Proof.* Necessary and sufficient conditions for the convergence of the series (11) with probability one were found by A. Ya. Khintchine and A. N. Kolmogorov [63] and consist in the following: For every $\epsilon>0$ the following three series should converge:

$$\sum_{k=1}^{\infty}\int_{|x|>\epsilon} dF_k(x),$$

$$\sum_{k=1}^{\infty}\int_{|x|<\epsilon} x\,dF_k(x),$$

$$\sum_{k=1}^{\infty}\Big\{\int_{|x|<\epsilon} x^2\,dF_k(x) - \Big(\int_{|x|<\epsilon} x\,dF_k(x)\Big)^2\Big\}.$$

For the convergence of these series it is necessary and sufficient that for every $m > n$

$$\sum_{k=n}^{m} \int_{|x|>\varepsilon} dF_k(x) \to 0 \qquad (n \to \infty),$$

$$\sum_{k=n}^{m} \int_{|x|<\varepsilon} x\, dF_k(x) \to 0 \qquad (n \to \infty),$$

$$\sum_{k=n}^{m} \left\{ \int_{|x|<\varepsilon} x^2 dF_k(x) - \Big( \int_{|x|<\varepsilon} x\, dF_k(x)\Big)^2 \right\} \to 0 \qquad (n \to \infty).$$

These relations prove our assertion by Remark 3 after Theorem 1.

This corollary may be formulated in another way. In order that the sums

$$\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n$$

of independent random variables should converge with probability one, it is necessary and sufficient that they converge in probability.

As another corollary of Theorem 3 we shall state a result obtained by A. Ya. Khintchine [51].

COROLLARY 2. *If the random variables*

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

*are independent, identically distributed, and have a finite mathematical expectation* $\mathbf{M}\xi_n = a$, *then the law of large numbers applies to them,* i.e., *for every* $\epsilon > 0$ *and* $n \to \infty$

$$\mathsf{P}\left\{\left| \frac{1}{n}\sum_{k=1}^{n} \xi_k - a \right| < \varepsilon\right\} \to 0.$$

*Proof.* For the proof it is sufficient to verify that in the case considered all the conditions of Theorem 3 are satisfied. To this end we note that, by hypothesis,

$$\int x\, dF(x+a) = 0 \tag{12}$$

and

$$\int |x|\, dF(x+a) < +\infty. \tag{13}$$

Now,

$$\sum_{k=1}^{n} \int_{|x|>n} dF_k(x+a_k) = n \int_{|x|>n} dF(x+a) \leqslant \int_{|x|>n} |x|\, dF(x+a),$$

so that by (13) the first condition of Theorem 3 is satisfied. Furthermore,

$$\sum_{k=1}^{n} \frac{1}{n} \int_{|x|<n} x\, dF_k(x+a_k) = \int_{|x|<n} x\, dF(x+a),$$

so that by (12) the second condition of Theorem 3 is satisfied. Let $L > 0$ be arbitrary. For sufficiently large $n$, we have

$$\sum_{k=1}^{n} \frac{1}{n^2} \int_{|x|<n} x^2 \, dF_k(x + a_k) = \frac{1}{n} \int_{|x|<n} x^2 \, dF(x+a)$$
$$\leqslant \frac{1}{n} \int_{|x|<L} x^2 \, dF(x+a) + \int_{L \leqslant |x| < n} |x| \, dF(x+a).$$

Whatever the constant $L$ may be, the first term in the last sum converges to zero. The second term can be made arbitrarily small by properly choosing $L$. Khintchine's theorem is thereby proved.

## § 28. Relative Stability

The concept of relative stability of the sums of a sequence $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ of positive random variables was introduced by A. Ya. Khintchine [54]. Namely, the sums

$$\zeta_n = \xi_1 + \xi_2 + \cdots + \xi_n$$

of positive random variables $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ are said to be *relatively stable*, if it is possible to find constants $B_n > 0$ such that for every $\epsilon > 0$

$$\mathsf{P}\left\{\left|\frac{\zeta_n}{B_n} - 1\right| > \varepsilon\right\} \to 0 \quad (n \to \infty).$$

For the scheme of a double sequence we shall say that the sums

$$\zeta_n = \xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n} \tag{1}$$

of positive random variables are *relatively stable*, if for every $\epsilon > 0$

$$\mathsf{P}\{|\xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n} - 1| > \varepsilon\} \to 0 \qquad (n \to \infty).$$

Clearly, relative stability of the sums is the most natural and general form of the law of large numbers for positive random variables. Hence * we conclude that if the sums (1) are relatively stable, then the separate summands are asymptotically constant.† This circumstance permits us ‡ to formulate the following theorem:

THEOREM 1.§ *In order that the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n}$$

* *Translator's note.* From Theorem 1 of § 27 or 1) below.

† The summands are assumed to be independent.

‡ *Translator's note.* The authors probably mean that we need not explicitly assume that the variables are asymptotically constant. See below (8) of § 22.

§ B. V. Gnedenko [41].

*of independent positive random variables be relatively stable, it is necessary and sufficient that as* $n \to \infty$

$$1)\quad \sum_{k=1}^{k_n} \int \frac{x^2}{1+x^2}\, dF_{nk}(x+m_{nk}) \to 0,$$

$$2)\quad \sum_{k=1}^{k_n} \Big\{ m_{nk} + \int_{|x|<1} x\, dF_{nk}(x+m_{nk}) \Big\} \to 1.$$

*Proof.* The theorem is an immediate consequence of Theorem 1, § 27.

In the following theorems aiming at establishing the connections between the conditions for convergence to the normal law and the conditions for the relative stability of the sums, we shall assume that the separate summands $\xi_{nk}$ are infinitesimal.

THEOREM 2. *In order that the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n}$$

*of independent positive random variables be relatively stable and the variables* $\xi_{nk}$ *be infinitesimal, it is necessary and sufficient that for every* $\epsilon > 0$

$$1)\quad \sum_{k=1}^{k_n} \int_{\epsilon}^{\infty} dF_{nk}(x) \to 0 \qquad (n \to \infty),$$

$$2)\quad \sum_{k=1}^{k_n} \int_{0}^{\epsilon} x\, dF_{nk}(x) \to 1 \qquad (n \to \infty).$$

*Proof.* To prove the theorem it is sufficient to show that under our conditions the third relation in Remark 3 after Theorem 1 of § 27 is a consequence of the first two. This follows since for every $\delta$ $(0 < \delta < \epsilon)$ we have

$$0 \leqslant \sum_{k=1}^{k_n} \Big\{ \int_0^{\epsilon} x^2\, dF_{nk}(x) - \Big( \int_0^{\epsilon} x\, dF_{nk}(x) \Big)^2 \Big\}$$

$$\leqslant \sum_{k=1}^{k_n} \int_0^{\epsilon} x^2\, dF_{nk}(x) \leqslant \delta \sum_{k=1}^{k_n} \int_0^{\delta} x\, dF_{nk}(x) + \epsilon^2 \sum_{k=1}^{k_n} \int_{\delta}^{\infty} dF_{nk}(x).$$

The following two propositions follow readily from the theorem just proved.

COROLLARY 1. *If the random variables* $\xi_{nk}$ *are independent in each row and have mathematical expectations with* $\sum_{k=1}^{k_n} \mathbf{M}\xi_{nk} = 1$, *then in order that the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n}$$

*be relatively stable and the variables $\xi_{nk}$ be infinitesimal, it is necessary and sufficient that for every $\epsilon > 0$*

$$\sum_{k=1}^{k_n} \int_{\epsilon}^{\infty} x\, dF_{nk}(x) \to 0 \qquad (n \to \infty).$$

In particular, we obtain from this

COROLLARY 2. *If the positive independent random variables $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ have mathematical expectations, then in order that the sums*

$$\zeta_n = \xi_1 + \xi_2 + \cdots + \xi_n$$

*be relatively stable for the particular coefficients $B_n$,*

$$B_n = \sum_{k=1}^{n} \mathbf{M}\,\xi_k$$

*and the variables $\xi_k/B_n$, $(1 \leq k \leq n)$ be infinitesimal, it is necessary and sufficient that for every $\epsilon > 0$*

$$\sum_{k=1}^{n} \frac{1}{B_n} \int_{\epsilon B_n}^{\infty} x\, dF_k(x) \to 0 \qquad (n \to \infty).$$

THEOREM 3.* *In order that for the sequence $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ of positive independent random variables the sums*

$$\zeta_n = \xi_1 + \xi_2 + \cdots + \xi_n$$

*be relatively stable and the variables $\xi_k/B_n$, $(1 \leq k \leq n)$ be infinitesimal for suitably chosen constants $B_n$, it is necessary and sufficient that there exist a sequence of positive constants $C_1, C_2, \ldots, C_n, \ldots$ such that*

1) $$\sum_{k=1}^{n} \int_{C_n}^{\infty} dF_k(x) \to 0 \qquad (n \to \infty),$$

2) $$\sum_{k=1}^{n} \frac{1}{C_n} \int_{0}^{C_n} x\, dF_k(x) \to \infty \qquad (n \to \infty).$$

*Proof.* By Theorem 2 for the relative stability of the sums $\zeta_n = \xi_1 + \xi_2 + \cdots + \xi_n$ it is necessary that for every $\epsilon > 0$

1) $$\sum_{k=1}^{n} \int_{\epsilon}^{\infty} dF_k(B_n x) \to 0 \qquad (n \to \infty),$$

2) $$\sum_{k=1}^{n} \int_{0}^{\epsilon} x\, dF_k(B_n x) \to 1 \qquad (n \to \infty).$$

---

* A. A. Bobrov [12].

Obviously, it is possible to pick a sequence $\epsilon_n \to 0$ such that

$$\sum_{k=1}^{n} \int_{\epsilon_n}^{\infty} dF_k(B_n x) = \sum_{k=1}^{n} \int_{\epsilon_n B_n}^{\infty} dF_k(x) \to 0 \qquad (n \to \infty),$$

$$\sum_{k=1}^{n} \int_{0}^{\epsilon_n} x\, dF_k(B_n x) = \sum_{k=1}^{n} \frac{1}{B_n} \int_{0}^{\epsilon_n B_n} x\, dF_k(x) \to 1 \qquad (n \to \infty).$$

In order to complete the proof of the necessity of the conditions of the theorem it remains only to put $C_n = \epsilon_n B_n$.

Now let the conditions of the theorem be satisfied. Define $B_n$ by the formula

$$B_n = \sum_{k=1}^{n} \int_{0}^{C_n} x\, dF_k(x).$$

Then it follows from the second condition of the theorem that

$$C_n = o(B_n).$$

Consequently, for every $\epsilon > 0$ and sufficiently large $n$,

$$\sum_{k=1}^{n} \int_{\epsilon B_n}^{\infty} dF_k(x) \leqslant \sum_{k=1}^{n} \int_{C_n}^{\infty} dF_k(x).$$

Hence according to the first condition of the theorem it follows that as $n \to \infty$

$$\sum_{k=1}^{n} \int_{\epsilon B_n}^{\infty} dF_k(x) \to 0. \tag{2}$$

Furthermore, for sufficiently large $n$,

$$\sum_{k=1}^{n} \frac{1}{B_n} \int_{0}^{\epsilon B_n} x\, dF_k(x) = \sum_{k=1}^{n} \frac{1}{B_n} \int_{0}^{C_n} x\, dF_k(x) + \sum_{k=1}^{n} \frac{1}{B_n} \int_{C_n}^{\epsilon B_n} x\, dF_k(x)$$

$$= 1 + \sum_{k=1}^{n} \frac{1}{B_n} \int_{C_n}^{\epsilon B_n} x\, dF_k(x).$$

But by the first condition of the theorem, as $n \to \infty$

$$\sum_{k=1}^{n} \frac{1}{B_n} \int_{C_n}^{\epsilon B_n} x\, dF_k(x) \leqslant \varepsilon \sum_{k=1}^{n} \int_{C_n}^{\infty} dF_k(x) \to 0.$$

Thus for every $\epsilon > 0$

$$\frac{1}{B_n}\sum_{k=1}^{n}\int_{0}^{\epsilon B_n} x\,dF_k(x) \to 1 \tag{3}$$

as $n \to \infty$. According to the preceding theorem, it follows from (2) and (3) that the sums $\xi_1 + \xi_2 + \cdots + \xi_n$ are relatively stable and the variables $\xi_k/B_n$ are infinitesimal.

We shall now show the close connection between relative stability and convergence to the normal law.

THEOREM 4.* *In order that the distribution laws of the sums*

$$\zeta_n = \sum_{k=1}^{k_n} (\xi_{nk} - \mathsf{M}\,\xi_{nk})$$

*of random variables which are independent in each row and subject to the conditions*

$$\sum_{k=1}^{k_n} \mathsf{D}^2\,\xi_{nk} = 1, \quad \sup_{1 \leqslant k \leqslant k_n} \mathsf{P}\{|\xi_{nk} - \mathsf{M}\,\xi_{nk}| > \varepsilon\} \to 0$$

*for every $\epsilon > 0$, converge to the normal law*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt, \tag{4}$$

*it is necessary and sufficient that the sums*

$$\eta_n^2 = \sum_{k=1}^{k_n} (\xi_{nk} - \mathsf{M}\,\xi_{nk})^2$$

*of the squared deviations of the random variables from their mathematical expectations be relatively stable.*

If the existence of moments is not assumed, then the theorem we are interested in can be formulated as follows [37]:

THEOREM 5. *In order that for some suitably chosen constants $A_n$ the distribution laws of the sums*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \cdots + \xi_{nk_n} - A_n$$

*of independent infinitesimal random variables converge to the normal law (4), it is necessary and sufficient that the sums*

$$\eta_n^2 = \sum_{k=1}^{k_n}\Big(\xi_{nk} - \int_{|x|<1} x\,dF_{nk}(x)\Big)^2$$

---

* D. A. Raikov [89].

*of the squared deviations of the $\xi_{nk}$ from their "truncated mathematical expectations"* $\int\limits_{|x|<1} x\, dF_{nk}(x)$ *be relatively stable.*

*Proof.* If the distribution function of the random variable $\xi_{nk}$ is $F_{nk}(x)$, then that of the variables

$$\xi'_{nk} = \xi_{nk} - \int\limits_{|x|<1} x\, dF_{nk}(x) \quad \left[\text{In Theorem 4 } \xi'_{nk} = \xi_{nk} - \mathbf{M}\xi_{nk}\right]$$

is

$$F'_{nk}(x) = F_{nk}\Big(x + \int\limits_{|x|<1} x\, dF_{nk}(x)\Big)$$

[in Theorem 4 $F'_{nk}(x) = F_{nk}(x + \mathbf{M}\xi_{nk})$].

The distribution function of the random variable $\xi'^2_{nk}$ is determined at its continuity points by the equation

$$H_{nk}(y) = \mathsf{P}\{\xi'^2_{nk} < y\} = \mathsf{P}\{|\xi'_{nk}| < \sqrt{y}\} = F'_{nk}(\sqrt{y}) - F'_{nk}(-\sqrt{y}).$$

Now it is obvious that

$$1)\ \sum_{k=1}^{k_n} \int\limits_{|x|\geqslant\varepsilon} dF'_{nk}(x) = \sum_{k=1}^{k_n} \int\limits_{x\geqslant\varepsilon} d\{F'_{nk}(x) - F'_{nk}(-x)\}$$

$$= \sum_{k=1}^{k_n} \int\limits_{y\geqslant\varepsilon^2} d\{F'_{nk}(\sqrt{y}) - F'_{nk}(-\sqrt{y})\} = \sum_{k=1}^{k_n} \int\limits_{y\geqslant\varepsilon^2} dH_{nk}(y),$$

$$2)\ \sum_{k=1}^{k_n} \int\limits_{|x|<\varepsilon} x^2\, dF'_{nk}(x) = \sum_{k=1}^{k_n} \int\limits_{0\leqslant x<\varepsilon} x^2\, d\{F'_{nk}(x) - F'_{nk}(-x)\}$$

$$= \sum_{k=1}^{k_n} \int\limits_{0\leqslant y<\varepsilon^2} y\, dH_{nk}(y),$$

$$3)\ \sum_{k=1}^{k_n} \int\limits_{|x|>\varepsilon} x^2\, dF'_{nk}(x) = \sum_{k=1}^{k_n} \int\limits_{y\geqslant\varepsilon^2} y\, dH_{nk}(y).$$

The first two of these equations show that the conditions of Theorem 2 of this section and those of Theorem 2 of § 26 are either simultaneously satisfied or simultaneously violated. The last equation shows that the same is true of the first corollary to Theorem 2 and the particular case of Theorem 3 of § 21. The proof of Theorems 4 and 5 is thereby completed.

It is clear that analogous theorems can be formulated and proved for asymptotically constant variables.

## CHAPTER 6

# LIMIT THEOREMS FOR CUMULATIVE SUMS

### § 29. Distributions of the Class $L$

As we have already pointed out, the general statement of the problem concerning limit distributions for sums of independent random variables, considered in the last two chapters, belongs to the last two decades. In the classical investigations only two particular cases of this problem were considered, namely, one sought for the conditions which must be imposed on a *sequence* of random variables

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots \tag{1}$$

so that 1) the law of large numbers should hold, 2) the central limit theorem should hold. In 1936 A. Ya. Khintchine formulated for the classical scheme of a sequence of mutually independent random variables the general problem of determining the class of distributions which can appear as limits of the distributions of the sums

$$\zeta_n = \frac{1}{B_n} \sum_{k=1}^{n} \xi_k - A_n \tag{2}$$

for suitably chosen real constants $B_n > 0$ and $A_n$.

Just as in the general case, for the solution of this problem it is necessary to introduce reasonable restrictions; namely, we assume that the variables

$$\xi_{nk} = \frac{\xi_k}{B_n} \qquad (1 \leqslant k \leqslant n; \quad n = 1, 2, \ldots)$$

are *asymptotically constant.*

It is clear that under this assumption every limit distribution for the normalized sums (2) is necessarily infinitely divisible. However, the converse is not true: there exist infinitely divisible distributions which cannot be the limiting distributions of the sums (2) for any choice of the constants $B_n > 0$ and $A_n$ and any choice of the sequence (1). This circumstance makes it transparently clear why in the classical investigations in order to obtain the Poisson law as "the law of rare events" it was necessary to have recourse to the scheme of a double sequence, already considered above.

Following Khintchine, we shall say that the distribution function $F(x)$ *belongs to the class* $L$, if it is possible to find a sequence of independent random variables (1) such that for suitably chosen constants $B_n > 0$ and $A_n$ the distribution functions of the sums (2) converge to $F(x)$, and the variables $\xi_{nk} = \xi_k/B_n$, $(1 \leq k \leq n)$ are asymptotically constant.

We note that without loss of generality we may suppose the variables $\xi_{nk} = \xi_k/B_n$ to be infinitesimal in the following. In fact, if the $\xi_{nk}$ are asymptotically constant, then putting

$$\xi'_{nk} = \xi_{nk} - m_{nk} = \frac{\xi_k - m_k}{B_n}$$

($m_{nk}$ and $m_k$ are medians corresponding to $\xi_{nk}$ and $\xi_k$) and $A'_n = A_n - \sum_{k=1}^{n} \frac{m_k}{B_n}$, we see that the class of limit distributions of the sums (2) of asymptotically constant summands $\xi_{nk} = \xi_k/B_n$ coincides with the class of limit distributions of the sums (2) of infinitesimal summands

$$\xi'_{nk} = \frac{\xi_k - m_k}{B_n}.$$

P. Lévy [76] gave a complete characterization of the class $L$ in answer to a question raised by A. Ya. Khintchine.

Before turning to an exposition of P. Lévy's theorem, we pause to prove a lemma (Khintchine [59]).

LEMMA. *If the distribution function which is the limit of the distribution functions of the sums (2) of independent infinitesimal summands $\xi_{nk} = \xi_k/B_n$ is proper, then as $n \to \infty$*

(a) $$B_n \to \infty,$$

(b) $$\frac{B_{n+1}}{B_n} \to 1.$$

*Proof.* (a) We suppose that there exists a sequence of indices $n_1 < n_2 < \cdots < n_k < \cdots$ such that the numbers $B_{n_k}$ remain bounded. Without loss of generality, we may suppose the indices to have been chosen so that the $B_{n_k}$ converge to some number $B \neq \infty$ as $k \to \infty$. Let $t$ be any given number, then the numbers $t_k = tB_{n_k}$ converge to $tB$ as $k \to \infty$. By hypothesis, the variables $\xi_{ns}$ $(1 \leq s \leq n)$ are infinitesimal, and so as $k \to \infty$

$$f_s(t) = f_s\left(\frac{t_k}{B_{n_k}}\right) \to 1$$

uniformly in $s$ $(1 \leq s \leq n_k)$, i.e., for every $t$

$$f_s(t) = 1 \qquad (s = 1, 2, \ldots).$$

It follows readily that for every $t$,

$$f(t) = \lim_{k \to \infty} \prod_{s=1}^{n_k} f_s\left(\frac{t}{B_{n_k}}\right) = 1.$$

But this equation means that $F(x)$ is an improper distribution function. We have arrived at a contradiction, proving the first part of the lemma.

(b) We note that since the summands $\xi_{nk}$ are infinitesimal, the distribution functions of the sums

$$\frac{\xi_1 + \xi_2 + \dots + \xi_n}{B_{n+1}} - A_{n+1} \tag{3}$$

also converge to $F(x)$. If we denote by $F_n(x)$ the distribution function of the sum (2), then the distribution function of the sum (3) is equal to

$$F_n(B'_n x + A'_n), \text{ where } B'_n = \frac{B_{n+1}}{B_n}, \quad A'_n = \frac{B_{n+1}}{B_n} A_{n+1} - A_n.$$

According to Theorem 2 of § 10 it follows from this that $(B_{n+1}/B_n) \to 1$ as $n \to \infty$. The lemma is completely proved.

THEOREM 1.* *In order that the distribution function $F(x)$ belong to the class L, it is necessary and sufficient that for every $\alpha$ $(0 < \alpha < 1)$ $F(x)$ be the composition of $F(x/\alpha)$ and some other distribution function $F_\alpha(x)$.*

*Proof. Sufficiency.* By hypothesis, for every $\alpha$ $(0 < \alpha < 1)$

$$f(t) = f(\alpha t) f_\alpha(t),$$

where $f_\alpha(t)$ is some characteristic function. We note first of all that a function $f(t)$ satisfying this condition never vanishes. In fact, suppose for example that $f(2a) = 0$ and $f(t) \neq 0$ for $0 \leq t < 2a$. Then

$$1 = 1 - |f_\alpha(2a)|^2 \leqslant 4\{1 - |f_\alpha(a)|^2\} \tag{4}$$

for every $\alpha$ $(0 < \alpha < 1)$. But since the function $f(t)$ is continuous, as $\alpha \to 1$

$$f_\alpha(a) = \frac{f(a)}{f(\alpha a)} \to 1.$$

Thus the inequality (4) as $\alpha$ approaches one leads to a contradiction.

We construct independent random variables $\xi_k$ with the characteristic functions

$$f_{\frac{k-1}{k}}(kt) = \frac{f(kt)}{f((k-1)t)}.$$

The characteristic function of the sum

$$\zeta_n = \frac{1}{n} \sum_{k=1}^{n} \xi_k$$

is equal to

$$\prod_{k=1}^{n} f_{\frac{k-1}{k}}\left(\frac{k}{n} t\right) = \prod_{k=1}^{n} \frac{f\left(\frac{k}{n} t\right)}{f\left(\frac{k-1}{n} t\right)} = f(t).$$

Since the function $f(t)$ is continuous and never vanishes, it is evident that

$$f_{\frac{k-1}{k}}\left(\frac{k}{n} t\right) \Rightarrow 1 \qquad (n \to \infty).$$

---

* P. Lévy [76].

uniformly in $k$ ($1 \leq k \leq n$). Thus it is proved that $F(x)$ belongs to the class $L$.

*Necessity.* Now suppose that the sequence of independent random variables

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

is such that for some suitably chosen constants $B_n > 0$ and $A_n$ the distribution functions of the sums

$$\zeta_n = \frac{1}{B_n}\sum_{k=1}^{n} \xi_k - A_k$$

converge to a limit distribution function $F(x)$ and the variables $\xi_{nk} = \xi_k/B_n$ ($1 \leq k \leq n$) are infinitesimal. If $F(x)$ is an improper distribution function then the condition of the theorem is trivially verified and the distribution function $F_\alpha(x)$ will also be improper. Only the case that $F(x)$ is a proper distribution function requires a proof.

Expressing our hypothesis in terms of characteristic functions, we obtain: as $n \to \infty$

$$f_{\zeta_n}(t) = e^{-iA_n t}\prod_{k=1}^{n} f_k\left(\frac{t}{B_n}\right) \Rightarrow f(t), \tag{5}$$

$f_k\left(\frac{t}{B_n}\right) \Rightarrow 1$ uniformly in $k$ ($1 \leq k \leq n$), $|f(t)| \not\equiv 1$, $f(t) \neq 0$ for any $t$ (being a characteristic function of an infinitely divisible law).

According to the preceding lemma, for every given $\alpha$ ($0 < \alpha < 1$) it is possible to pick $m = m(n)$ ($m < n$) so that as $n \to \infty$

$$\frac{B_m}{B_n} \to \alpha. \tag{6}$$

We write $f_{\zeta_n}(t)$ in the following form:

$$f_{\zeta_n}(t) = \left[e^{-iA_m\alpha t}\prod_{k=1}^{m} f_k\left(\frac{t}{B_n}\right)\right]\cdot\left[e^{-it(A_n - A_m\alpha)}\prod_{k=m+1}^{n} f_k\left(\frac{t}{B_n}\right)\right]. \tag{7}$$

But by (5), as $m \to \infty$

$$e^{-iA_m t}\prod_{k=1}^{m} f_k\left(\frac{t}{B_m}\right) \Rightarrow f(t);$$

hence, because of (6),

$$e^{-iA_m\alpha t}\prod_{k=1}^{m} f_k\left(\frac{t}{B_n}\right) \Rightarrow f(\alpha t). \tag{8}$$

The relations (5) and (8), together with Theorem 2 of § 13, permit us to conclude that the second factor on the right side of (7) must approach some characteristic function $f_\alpha(t)$.

Thus we find in the limit that for every $\alpha$ $(0 < \alpha < 1)$

$$f(t) = f(\alpha t) f_\alpha(t).$$

Q.E.D.

For later purposes it is important to note that $f_\alpha(t)$ is the characteristic function of an infinitely divisible distribution. Indeed, $f_\alpha(t)$ is the limit of a sequence of characteristic functions of sums of independent and asymptotically constant random variables.

## § 30. Canonical Representation of Distributions of the Class $L$

Each distribution of the class $L$ is infinitely divisible; hence the logarithm of its characteristic function can be represented by Lévy's formula

$$\log f(t) = i\gamma t - \frac{\sigma^2 t^2}{2} + \int_{-\infty}^{0} \left\{ e^{itu} - 1 - \frac{itu}{1+u^2} \right\} dM(u)$$
$$+ \int_{0}^{\infty} \left\{ e^{itu} - 1 - \frac{itu}{1+u^2} \right\} dN(u). \tag{1}$$

The question naturally arises as to what special properties the functions $M(u)$ and $N(u)$ must possess in order that $f(t)$ be the characteristic function of a distribution of the class $L$. To this question a complete answer is given by the following theorem, discovered by P. Lévy [76].

Theorem 1. *In order that the distribution function $F(x)$ belong to the class $L$ it is necessary and sufficient that the functions $M(u)$ and $N(u)$ in the formula* (1) *have right and left derivatives for every value $u$ and that the functions*

$$uM'(u) \qquad (u < 0),$$
$$uN'(u) \qquad (u > 0)$$

*be nonincreasing [here $M'(u)$ and $N'(u)$ denote either the right or the left derivative, possibly different ones at different points].*

*Proof.* Let $f(t)$ be the characteristic function of a distribution belonging to the class $L$, and $\alpha$ an arbitrary number between 0 and 1. We find from (1) that

$$\log f(\alpha t) = i\gamma\alpha t - \frac{\sigma^2\alpha^2 t^2}{2} + \int_{-\infty}^{0} \left\{ e^{i\alpha ut} - 1 - \frac{i\alpha ut}{1+u^2} \right\} dM(u)$$
$$+ \int_{0}^{+\infty} \left\{ \qquad \right\} dN(u) = i\gamma_1 t - \frac{\sigma^2 t^2 \alpha^2}{2}$$
$$+ \int_{-\infty}^{0} \left\{ e^{iut} - 1 - \frac{iut}{1+u^2} \right\} dM\left(\frac{u}{\alpha}\right) + \int_{0}^{\infty} \left\{ \qquad \right\} dN\left(\frac{u}{\alpha}\right),$$

where

$$\gamma_1 = \alpha\gamma + \alpha \int_{-\infty}^{0} \frac{u^3(1-\alpha^2)}{(1+u^2)(1+\alpha^2u^2)}\, dM(u) + \alpha \int_{0}^{+\infty} \frac{u^3(1-\alpha^2)}{(1+u^2)(1+\alpha^2u^2)}\, dN(u).$$

Now

$$\log \frac{f(t)}{f(\alpha t)} = i\gamma_2 t - \frac{\sigma^2(1-\alpha^2)t^2}{2} + \int_{-\infty}^{0} \Big\{ \qquad \Big\} d\left(M(u) - M\left(\frac{u}{\alpha}\right)\right)$$

$$+ \int_{0}^{+\infty} \left\{ e^{itu} - 1 - \frac{itu}{1+u^2} \right\} d\left(N(u) - N\left(\frac{u}{\alpha}\right)\right). \tag{2}$$

By the results of the preceding theorem, $f(t)/f(\alpha t)$ is an infinitely divisible characteristic function; (2) gives its canonical representation. Hence we conclude that the functions $M(u) - M(u/\alpha)$ and $N(u) - N(u/\alpha)$ must be nondecreasing.† Therefore, whatever $u_1 < u_2 < 0$ and $0 < v_1 < v_2$ are, we must have

$$M(u_1) - M\left(\frac{u_1}{\alpha}\right) \leqslant M(u_2) - M\left(\frac{u_2}{\alpha}\right),$$
$$N(v_1) - N\left(\frac{v_1}{\alpha}\right) \leqslant N(v_2) - N\left(\frac{v_2}{\alpha}\right),$$

and so

$$\left.\begin{aligned} M\left(\frac{u_2}{\alpha}\right) - M\left(\frac{u_1}{\alpha}\right) &\leqslant M(u_2) - M(u_1), \\ N\left(\frac{v_2}{\alpha}\right) - N\left(\frac{v_1}{\alpha}\right) &\leqslant N(v_2) - N(v_1). \end{aligned}\right\} \tag{3}$$

Conversely, if the functions $M(u)$ and $N(u)$ satisfy the inequalities (3) for every $\alpha$ $(0 < \alpha < 1)$, then the functions $M(u) - M(u/\alpha)$ and $N(u) - N(u/\alpha)$ will be nondecreasing, and consequently $f(t)/f(\alpha t)$ will be the characteristic function of an infinitely divisible law. Therefore *the conditions* (3) *are necessary and sufficient for the distribution to belong to the class L.*

We confine our further discussion to the function $N(u)$, since it is possible to obtain similar results for $M(u)$ by the same arguments.

Let $a < b$, $h > 0$ and $\alpha = e^{a-b}$. By (3) we have

$$N(e^{a+h}) - N(e^a) \geqslant N\left(\frac{e^{a+h}}{\alpha}\right) - N\left(\frac{e^a}{\alpha}\right) = N(e^{b+h}) - N(e^b).$$

If we denote

$$N(e^v) = S(v),$$

then it follows from the preceding inequality that the nondecreasing function $S(v)$ satisfies the inequality

$$S(a+h) - S(a) \geqslant S(b+h) - S(b) \quad (a < b). \tag{4}$$

† *Translator's note.* For this conclusion we need the Corollary to Theorem 1 of § 18.

Thus, the increment of $S(v)$ in intervals of given length $h$ can only decrease, when the interval shifts from left to right.

This circumstance permits us to conclude first of all that the function $S(v)$ is continuous at every point $v = v_0$.

We put now $b = a + h$. We find from (4) that

$$S(a+h) \geqslant \frac{S(a) + S(a+2h)}{2},$$

i.e., the continuous function $S(v)$ is concave.* Therefore it has finite left and right derivatives at every point; the value of the right derivative never exceeds that of the left derivative; and both derivatives are nonincreasing as $x$ increases (see Hardy [48], p. 91).

Thus $S'(v)$ is a nonincreasing function [$S'(v)$ does not necessarily remain the right or the left derivative throughout]. But

$$S(v) = N(e^v), \quad S'(v) = e^v N'(e^v),$$

hence, returning to the notation $u = e^v$, we find that the conditions of the theorem are necessary.

Now suppose that $uM'(u)$ and $uN'(u)$ are nonincreasing functions. Then, for every $\alpha$ $(0 < \alpha < 1)$,

$$\frac{u}{\alpha} M'\left(\frac{u}{\alpha}\right) \geqslant uM'(u) \qquad \text{for } u < 0,$$

$$\frac{u}{\alpha} N'\left(\frac{u}{\alpha}\right) \leqslant uN'(u) \qquad \text{for } u > 0.$$

Hence if $v_1 < v_2 < 0$, then

$$\int_{v_1}^{v_2} dM\left(\frac{v}{\alpha}\right) = \int_{v_1}^{v_2} \frac{1}{\alpha} M'\left(\frac{v}{\alpha}\right) dv \leqslant \int_{v_1}^{v_2} M'(v)\, dv = \int_{v_1}^{v_2} dM(v),$$

and if $0 < u_1 < u_2$, then

$$\int_{u_1}^{u_2} dN\left(\frac{u}{\alpha}\right) = \int_{u_1}^{u_2} \frac{1}{\alpha} N'\left(\frac{u}{\alpha}\right) du \leqslant \int_{v_1}^{v_2} N'(u)\, du = \int_{v_1}^{v_2} dN(u).$$

These inequalities, as we have seen above, prove that the distribution $F(x)$ belongs to the class $L$.

The simplest example of a distribution belonging to the class $L$ is the normal distribution. Then the functions $uM'(u)$ and $uN'(u)$ are identically equal to zero.

---

* *Translator's note.* In the original the word "convex" was written instead of "concave"; further on, the words "right" and "left" were interchanged and "increase" was written instead of "do not increase."

We now determine what conditions must be imposed on the function $K(u)$ in Kolmogorov's formula in order that a distribution with finite variance belong to the class $L$. Since the functions $M(u)$ and $N(u)$ are related to $K(u)$ by the formulas

$$dM(u) = \frac{1}{u^2}\, dK(u) \qquad \text{for } u < 0,$$

$$dN(u) = \frac{1}{u^2}\, dK(u) \qquad \text{for } u > 0,$$

the following result [46] is readily obtained from Theorem 1.

THEOREM 2. *In order that the distribution function $F(x)$ with finite variance belong to the class $L$, it is necessary and sufficient that the function $K(u)$ in Kolmogorov's formula have right and left derivatives at every point $u \neq 0$ and that the function $K'(u)\,u$, where $K'(u)$ denotes the right or left derivative, possibly different ones at different points, be nonincreasing for $u < 0$ and $u > 0$.*

It is easily verified that distributions with finite variances belonging to the class $L$ are not exhausted by the normal and the improper distributions. For example, the distribution for which

$$K(u) = \begin{cases} 0 & \text{for } u < 0, \\ u^2 & \text{for } 0 \leqslant u < 1, \\ 1 & \text{for } u \geqslant 1, \end{cases}$$

in Kolmogorov's formula, satisfies the conditions of Theorem 2 and so belongs to the class $L$.

## § 31. CONDITIONS FOR CONVERGENCE

In this section we shall derive conditions which it is necessary to impose on the distribution functions of independent random variables $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ in order that for suitably chosen constants $B_n > 0$ and $A_n$ the distribution functions of the sums

$$\zeta_n = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{B_n} - A_n \tag{1}$$

converge to some limit distribution function and the summands $\xi_{nk} = \dfrac{\xi_n}{B_n}$ $(1 \leq k \leq n)$ be asymptotically constant.*

* Throughout this section the case of an improper limit distribution is excluded.

If the constants $B_n$ are given in advance, then the solution of the problem posed above is already contained in the theorems of § 25. The essential difference between the present problem and the general problem solved before is that there is now indicated a general rule by which the constants $B_n$ should be chosen.

If the variables $\xi_n$ have finite variances and if conditions are sought for the convergence of the distribution functions and the variances of the sums (1) to a limit distribution function and its variance, then the question receives a most simple answer.

THEOREM 1.* *In order that for suitably chosen constants $A_n$ and $B_n > 0$ the distribution functions of the sums* (1) *converge to a limit, their variances converge to the variance of the limit, and the summands* $\frac{1}{B_n}(\xi_k - \mathbf{M}\xi_k)$ *be infinitesimal, it is necessary and sufficient that there exist a nondecreasing function $K_1(u)$ with variation equal to one such that for all $u \neq 0$*

$$\frac{1}{C_n^2}\sum_{k=1}^{n}\int_{-\infty}^{C_n u} x^2\, dF_k(x + \mathbf{M}\xi_k) \to K_1(u) \tag{2}$$

*as $n \to \infty$, and moreover that*

$$\sup_{1 \leqslant k \leqslant n}\int \frac{x^2}{C_n^2 + x^2}\, dF_k(x + \mathbf{M}\xi_k) \to 0, \tag{3}$$

*where*

$$C_n^2 = \sum_{k=1}^{n} \mathbf{D}^2\xi_k.$$

*The constants $A_n$ and $B_n$ may be chosen according to the formulas*

$$B_n^2 = \sum_{k=1}^{n}\mathbf{D}^2\xi_k,\quad A_n = \frac{1}{B_n}\sum_{k=1}^{n}\mathbf{M}\xi_k.$$

*The limit distribution is defined by Kolmogorov's formula with the constant $\gamma = 0$ and the function $K(u) = K_1(u)$.*

*Proof.* Let $\xi_{nk} = (\xi_k - \mathbf{M}\xi_k)/C_n$; then $F_{nk}(x) = F_k(C_n x + \mathbf{M}\xi_k)$, and by (4) of § 20 the condition

$$\sup_{1 \leqslant k \leqslant n}\int \frac{x^2}{C_n^2 + x^2}\, dF_k(x + \mathbf{M}\xi_k) = \sup_{1 \leqslant k \leqslant n}\int \frac{z^2}{1 + z^2}\, dF_{nk}(z) \to 0$$

simply means that the variables $\xi_{nk}$ $(1 \leq k \leq n)$ are infinitesimal.

---

* B. V. Gnedenko and A. V. Groshev [26].

Suppose that the conditions of the theorem are satisfied. Then the choice of $C_n$ implies that as $n \to \infty$

$$\sum_{k=1}^{n} \int_{-\infty}^{u} x^2 \, dF_{nk}(x) \Longrightarrow K_1(u).$$

According to Theorem 2 of § 21 the distribution functions of the sums

$$\zeta_n = \frac{\xi + \ldots + \xi_n}{C_n} - A_n$$

converge to a limit which is given by putting $K(u) = K_1(u)$ in Kolmogorov's formula. If we take $\gamma = 0$ in Kolmogorov's formula for the limit distribution then $A_n$ must be chosen as indicated in the statement of the theorem.

We shall now prove the necessity of the conditions of the theorem. Suppose that for some $A_n$ and $B_n > 0$ the distribution functions of the sums (1) and their variances converge to a limit distribution function and its variance and the summands $(\xi_k - \mathbf{M}\xi_k)/B_n$ are infinitesimal. According to Theorem 2 of § 21, under these conditions there exists a nondecreasing function $K(u)$ such that at all continuity points of $K(u)$, hence by Theorem 2 of § 30 at all $u \neq 0$ and also for $u = \infty$,

$$\sum_{k=1}^{n} \int_{-\infty}^{u} x^2 \, dF_k(B_n x + \mathbf{M}\,\xi_k) \to K(u) \quad (n \to \infty). \tag{4}$$

If $V > 0$ is the variation of $K(u)$, then we put $v = u\sqrt{V}$ and

$$K_1(v) = \frac{1}{V} K(u).$$

By (4), for all $v \neq 0$

$$\sum_{k=1}^{n} \int_{-\infty}^{u} \frac{x^2}{V} \, dF_k(B_n x + \mathbf{M}\,\xi_k) = \sum_{k=1}^{n} \int_{-\infty}^{v} z^2 \, dF_k(B_n z\sqrt{V} + \mathbf{M}\,\xi_k) \to K_1(v)$$

and, in particular, for $v = +\infty$

$$\frac{1}{VB_n^2} \sum_{k=1}^{n} \int (x - \mathbf{M}\,\xi_k)^2 \, dF_k(x) = \frac{1}{VB_n^2} \sum_{k=1}^{n} \mathbf{D}^2\,\xi_k \to K_1(+\infty) = 1.$$

By choosing new $B_n$ according to the formula

$$VB_n^2 = \sum_{k=1}^{n} \mathbf{D}^2\,\xi_k,$$

we do not change the limit law for the sums (1), as follows from Theorem 2 of § 10. Q.E.D.

We now turn to the consideration of the general case. Suppose that the random variables $\xi_k$ and $\xi_k'$ have the same distribution function $F_k(x)$ and that $\xi_k'$ is independent of $\xi_l$ for all $l$. The distribution function of the difference $\eta_k = \xi_k - \xi_k'$ is denoted by $V_k(x)$. Obviously, $V_k(x)$ is a symmetrical distribution function and its characteristic function $v_k(t)$ is $|f_k(t)|^2$. Now

suppose that for suitably chosen constants $B_n > 0$ and $A_n$ the distribution functions of the sums

$$\zeta_n = \frac{\xi_1 + \xi_2 + \dots + \xi_n}{B_n} - A_n \tag{5}$$

converge to a limit distribution function $F(x)$ as $n \to \infty$. Then it is obvious that the distribution functions of the sums

$$\frac{\eta_1 + \eta_2 + \dots + \eta_n}{B_n} = \left(\frac{\xi_1 + \xi_2 + \dots + \xi_k}{B_n} - A_n\right) - \left(\frac{\xi_1' + \xi_2' + \dots + \xi_n'}{B_n} - A_n\right) \tag{6}$$

converge to the function $V(x) = F(x) \ast [1 - F(-x + 0)]$, the characteristic function of which is equal to $v(t) = |f(t)|^2$.

THEOREM 2.† *In order that for suitably chosen constants $B_n > 0$ and $A_n$ the distribution functions of the sums* (5) *converge to a proper distribution function $F(x)$ and the summands $\xi_k/B_n$ $(1 \le k \le n)$ be infinitesimal, it is necessary and sufficient that there exist a nondecreasing function $G^*(u)$ $[G^*(-\infty) = 0]$ with finite variation $V = G^*(+\infty)$ such that, if we define $C_n > 0$ by the equation*

$$\sum_{k=1}^{n} \int \frac{x^2}{C_n^2 + x^2}\, dV_k(x) = 2V, \tag{7}$$

*then we have*

($\alpha$) *for every $u$ $(-\infty \le u \le +\infty)$ except possibly $u = 0$,*

$$\sum_{k=1}^{n} \int_{-\infty}^{C_n u} \frac{x^2}{C_n^2 + x^2}\, dF_k(x + \alpha_{nk}) \to G^*(u) \quad \text{as} \quad n \to \infty,$$

*where*

$$\alpha_{nk} = \frac{1}{C_n} \int_{|x| < C_n} x\, dF_k(x);$$

($\beta$) $$\sup_{1 \leqslant k \leqslant n} \int \frac{x^2}{C_n^2 + x^2}\, dF_k(x) \to 0 \quad \text{as} \quad n \to \infty.$$

*Proof.* The condition ($\beta$) means simply that the summands $\xi_k/C_n$, $(1 \le k \le n)$ are infinitesimal.

Since $\xi_{nk} = \xi_k/B_n$ and $F_{nk}(x) = F_k(B_n x)$, it follows from Theorem 1 of § 25 that a necessary and sufficient condition for the convergence of the distributions of the sums (5) to a limit is the following:

---

† B. V. Gnedenko and A. V. Groshev [46].

$$\lim_{n\to\infty}\sum_{k=1}^{n}\int_{-\infty}^{B_n u}\frac{x^2}{B_n^2+x^2}\,dF_k(x+\alpha_{nk})=G^*(u) \tag{8}$$
$$(-\infty\leqslant u\leqslant\infty).$$

The sufficiency of ($\alpha$) is now obvious. In fact, it means that (8) is satisfied with $B_n = C_n$.

Now suppose that the distribution functions of the sums (5) converge to $F(x)$ for some choice of the constants $B_n$. If we prove that $(B_n\ C_n) \to 1$ as $n \to \infty$, then by Theorem 2 of § 10 the distribution functions of the sums

$$\frac{\xi_1+\xi_2+\ldots+\xi_n}{C_n}-A_n,$$

also converge to $F(x)$, and so ($\alpha$) is satisfied in view of (8).

The distribution functions of the sums (6), as already stated, converge to $V(x)$. The function $G(u)$ in the formula of Lévy and Khintchine for $V(x)$ is

$$G(u)=G^*(u)+V-G^*(-u).$$

The median of the random variable $\xi_k$ is equal to zero. Therefore by Theorem 3 of § 25 for the convergence of the distribution functions of the sums (6) to $V(x)$ it is necessary that for $-\infty \leq u \leq +\infty$, except possibly for $u = 0$,

$$\sum_{k=1}^{n}\int_{-\infty}^{B_n u}\frac{x^2}{B_n^2+x^2}\,dV_k(x)\to G(u)\quad\text{as}\quad n\to\infty. \tag{9}$$

From (7) and (9) (for $u = +\infty$) we conclude that as $n \to \infty$

$$R_n=\sum_{k=1}^{n}\left\{\int\frac{x^2}{B_n^2+x^2}\,dV_k(x)-\int\frac{x^2}{C_n^2+x^2}\,dV_k(x)\right\}\to 0. \tag{10}$$

By hypothesis, $F(x)$ is a proper distribution function. This means that $G^*(u) \not\equiv 0$ and consequently $G(u) \not\equiv 0$. Therefore it is possible to find $a > 0$ such that

$$\Delta=G(a)-G(-a)>0.$$

From (9) we conclude that for $n > n_0$

$$\sum_{k=1}^{n}\int_{-B_n a}^{B_n a}\frac{x^2}{B_n^2+x^2}\,dV_k(x)>\frac{\Delta}{2}.$$

For $n > n_0$ we have the following chain of obvious inequalities:

$$|R_n| = |B_n^2 - C_n^2| \sum_{k=1}^{n} \int \frac{x^2}{(B_n^2 + x^2)(C_n^2 + x^2)} dV_k(x)$$

$$\geqslant \frac{|B_n^2 - C_n^2|}{(a^2 B_n^2 + C_n^2)} \sum_{k=1}^{n} \int_{-B_n a}^{B_n a} \frac{x^2}{B_n^2 + x^2} dV_k(x) > \frac{\left|\left(\frac{B_n}{C_n}\right)^2 - 1\right|}{\left|a^2 \frac{B_n^2}{C_n^2} + 1\right|} \cdot \frac{\Delta}{2} \geqslant 0.$$

By (10) it follows from this that $(B_n/C_n) \to 1$ as $n \to \infty$. Q.E.D.

## § 32.* Unimodality of Distributions of the Class $L$

In mathematical statistics a distribution function is called unimodal if its derivative $F'(x)$ exists everywhere and has a unique (finite) maximum. Following A. Ya. Khintchine, we generalize this concept somewhat.

Definition. The distribution function $F(x)$ is called *unimodal* if there exists at least one value $x = a$ such that $F(x)$ is convex for $x < a$ and concave for $x > a$.

It is easy to verify that the normal distribution, the Cauchy law $F(x) = \frac{1}{2} + (1/\pi) \arctan x)$, the uniform distribution in a finite interval, and the improper distribution $\epsilon(x)$ are all unimodal in the sense just described.

Without loss of generality, we shall suppose in the following theorems that the vertex of the distribution (the point $a$) is located at the point $a = 0$. It follows from the definition cited above that a unimodal distribution function has right and left derivatives at every point except possibly the vertex. Moreover, $F'(x)$ [under $F'(x)$ we mean either one of the derivatives — right or left — possibly different ones at different points] does not decrease for $x < a$ and does not increase for $x > a$. This fact is well known in the theory of convex functions (see Hardy [48], p. 91).

Theorem 1.‡ *In order that the distribution function $F(x)$ be unimodal (with vertex at $x = 0$), it is necessary and sufficient that the function*

$$V(x) = F(x) - xF'(x)$$

*be a distribution function.*§

*Proof.* First we make a few elementary remarks.

---

* *Translator's note.* § 32 has been shortened; see Appendix II.

‡ A. Ya. Khintchine [60].

§ *Translator's note.* See Appendix II for a discussion of the theorem. Part of the result was apparently rediscovered by N. L. Johnson and C. A. Rogers, The moment problem for unimodal distributions, *Annals of Mathematical Statistics*, 22, 433–439 (1951).

1°. If $F(x)$ is a unimodal distribution function, then for $x \to +0$, $x \to -0$, $x \to +\infty$, and $x \to -\infty$,

$$xF'(x) \to 0. \tag{1}$$

Indeed, for $x > 0$ [$F'(x)$ does not increase!]

$$xF'(x) \leqslant 2 \int_{\frac{x}{2}}^{x} F'(u)\, du.$$

Hence

$$0 \leqslant xF'(x) \leqslant 2\left(1 - F\left(\frac{x}{2}\right)\right) \to 0$$

as $x \to +\infty$ and

$$0 \leqslant xF'(x) \leqslant 2\,(F(x) - F(+0)) \to 0$$

as $x \to +0$. In exactly the same way (1) is proved for $x \to -0$ and $x \to -\infty$.

2°. If $V(x)$ is a distribution function, then as $x \to +0$ or $x \to +\infty$

$$x \int_{x}^{\infty} \frac{dV(u)}{u} \to 0. \tag{2a}$$

and as $x \to -0$ or $x \to -\infty$

$$x \int_{-\infty}^{x} \frac{dV(u)}{u} \to 0. \tag{2b}$$

In fact, as $x \to +0$

$$x \int_{x}^{\infty} \frac{dV(u)}{u} = x \int_{x}^{\sqrt{x}} \frac{dV(u)}{u} + x \int_{\sqrt{x}}^{\infty} \frac{dV(u)}{u}$$

$$\leqslant \int_{x}^{\sqrt{x}} dV(u) + \sqrt{x} \int_{\sqrt{x}}^{\infty} dV(u) \leqslant V(\sqrt{x}) - V(x) + \sqrt{x} \to 0.$$

As $x \to +\infty$

$$x \int_{x}^{\infty} \frac{dV(u)}{u} \leqslant 1 - V(x) \to 0.$$

In exactly the same way (2b) is proved.

Now let $F(x)$ be a unimodal distribution function. We shall prove that

$$V(x) = F(x) - xF'(x) \tag{3}$$

is a distribution function. Indeed, for $h > 0$ we have

$$\begin{aligned} V(x+h) - V(x) &= [F(x+h) - F(x) - hF'(x+h)] \\ &\quad + x[F'(x) - F'(x+h)] = [F(x+h) - F(x) - hF'(x)] \\ &\qquad + (x+h)[F'(x) - F'(x+h)]. \end{aligned}$$

For $x > 0$ in the first of these equations, and for $x < x + h < 0$ in the second, both summands on the right side are non-negative. Hence $V(x+h) - V(x) \geq 0$ for every $x \neq 0$. By (1) we have

$$V(+0) = F(+0), \qquad V(-0) = F(-0),$$
$$V(+\infty) = F(+\infty) = 1, \quad V(-\infty) = F(-\infty) = 0.$$

We shall now prove the converse proposition. Namely, we shall prove that if $V(x)$ is an arbitrary distribution and the distribution function $F(x)$ satisfies equation (3), then $F(x)$ is unimodal. First of all, it is easy to convince ourselves that among the solutions of equation (3) there can be only one distribution function. Indeed, if there were two, $F_1(x)$ and $F_2(x)$, then their difference $W(x)$ would satisfy the equation

$$W(x) - xW'(x) = 0$$

and so *

$$F_1(x) - F_2(x) = cx$$

for $x \neq 0$, where $c$ is a constant. From the conditions $F_1(-\infty) = F_2(-\infty) = 0$ and the conditions $F_1(+\infty) = F_2(+\infty) = 1$ we conclude that $c = 0$. It is not difficult to convince ourselves by some elementary calculations that the function $F(x)$ defined by means of the equations

$$\left.\begin{aligned} F(x) &= -\int_{-\infty}^{x} du \int_{-\infty}^{u} \frac{dV(v)}{v} \qquad \text{for } x < 0, \\ F(x) &= 1 - \int_{x}^{\infty} du \int_{u}^{\infty} \frac{dV(v)}{v} \qquad \text{for } x > 0 \end{aligned}\right\} \tag{4}$$

satisfies (3).† Moreover, it obviously does not decrease for $x < 0$ and $x > 0$; furthermore, by (2a)

$$F(+0) = 1 - \int_{+0}^{\infty} du \int_{u}^{\infty} \frac{dV(v)}{v}$$
$$= 1 - u \int_{u}^{\infty} \frac{dV(v)}{v} \Bigg|_{+0}^{\infty} - \int_{+0}^{\infty} dV(v) = V(+0)$$

and in exactly the same way

$$F(-0) = V(-0).$$

In other words, $F(x)$ is a distribution function. By differentiating (4) we find that $F'(x)$ does not decrease for $x < 0$ and does not increase for $x > 0$. This proves that the function $F(x)$ is unimodal.

---

* *Translator's note.* See Appendix II for a reference to the required theorem.

† *Translator's note.* In verifying (3) the denumerable set of points of discontinuity of $V(x)$ may be ignored; see Appendix II.

The theorem just proved can be formulated in terms of characteristic functions as follows:

THEOREM 2.* *The function $f(t)$ is the characteristic function of a unimodal distribution function if and only if it can be represented in the form*

$$f(t)=\frac{1}{t}\int_0^t v(u)\,du,$$

*where $v(u)$ is some characteristic function.*

*Proof.* We have by (4)

$$f(t)=\int e^{itx}\,dF(x)$$

$$=-\int_{-\infty}^{-0} e^{itx}\left(\int_{-\infty}^{x}\frac{dV(v)}{v}\right)dx+F(+0)-F(-0)+\int_{+0}^{\infty} e^{itx}\left(\int_{x}^{\infty}\frac{dV(v)}{v}\right)dx.$$

But

$$\int_{+0}^{\infty} e^{itx}\left(\int_{x}^{\infty}\frac{dV(v)}{v}\right)dx=\int_{+0}^{\infty}\frac{e^{itx}-1}{it}\frac{dV(x)}{x}=\frac{1}{t}\int_{+0}^{\infty}\left\{\int_0^t e^{iux}\,du\right\}dV(x),$$

$$-\int_{-0}^{\infty} e^{itx}\int_{-\infty}^{x}\frac{dV(v)}{v}\,dx=\frac{1}{t}\int_{-\infty}^{-0}\left\{\int_0^t e^{iux}\,du\right\}\,dV(x),$$

$$F(+0)-F(-0)=\left(\frac{1}{t}\int_0^t e^{iu0}\,du\right)\cdot(V(+0)-V(-0)).$$

Thus

$$f(t)=\frac{1}{t}\int\left\{\int_0^t e^{iux}\,du\right\}dV(x)=\frac{1}{t}\int_0^t v(u)\,du.$$

THEOREM 3. *The composition of unimodal distribution functions is unimodal.*†

THEOREM 4.‡ *If a sequence of unimodal distribution functions converges to a distribution function, then the limit function is also unimodal.*

*Proof.* By hypothesis, as $n\to\infty$

$$F_n(x)\Rightarrow F(x).$$

Let $a_n$ be the vertex of $F_n(x)$ and $a=\overline{\lim\limits_{n\to\infty}}\,a_n$. Pick a subsequence of the indices $n_k$ so that $\lim\limits_{k\to\infty} a_{n_k}=a$. Now take any two continuity points of

* A. Ya. Khintchine [60].

† *Translator's note.* For a discussion of this incorrect statement, see Appendix II.

‡ A. I. Lapin [71].

$F(x)$, $x_1 < a$ and $x_2 < a$, and determine $N$ so that for all $k > N$ the inequalities $x_1 < a_{n_k}$ and $x_2 < a_{n_k}$ hold. By the hypothesis of unimodality of the functions $F_n(x)$, we have for $k > N$

$$F_{n_k}(x_1) + F_{n_k}(x_2) \geqslant 2F_{n_k}\left(\frac{x_1 + x_2}{2}\right).$$

This relation becomes, in the limit,

$$F(x_1) + F(x_2) \geqslant 2F\left(\frac{x_1 + x_2}{2}\right).$$

We have supposed $x_1$ and $x_2$ to be continuity points of $F(x)$; however, it is obvious that we can easily drop this restriction by making use of the continuity of $F(x)$ from the left.

In exactly the same way, for any $x_3 > a$ and $x_4 > a$ we find that

$$F(x_3) + F(x_4) \leqslant 2F\left(\frac{x_3 + x_4}{2}\right).$$

Thus, $F(x)$ is concave for $x > a$ and convex for $x < a$; namely, $F(x)$ is unimodal.

Finally we remark that $a \neq \pm\infty$, since otherwise $F(x)$ would be convex or concave for all values of the argument. For functions of bounded variation this is possible only if $F(x)$ is a constant.

# *Part III* IDENTICALLY DISTRIBUTED SUMMANDS

## CHAPTER 7

### § 33. Statement of the Problem, Stable Laws

We shall now turn to the detailed study of limit distributions of normalized sums

$$\zeta_n = \frac{\xi_1 + \xi_2 + \dots + \xi_n}{B_n} - A_n \tag{1}$$

of independent, *identically distributed* * random variables $\xi_1, \xi_2, \dots, \xi_n, \dots$.

By Theorem 4 of § 14, if the distributions of the sums (1) converge to a limit, then the variables $\xi_{nk} = (\xi_k/B_n) - (A_n/n)$ must necessarily be infinitesimal. This circumstance permits us to conclude that every distribution which is a limit distribution of the sums (1) must belong to the class $L$. The problem naturally arises of determining all possible limit distributions of the sums (1). The solution of this problem requires the introduction of a new concept.

Definition. The distribution function $F(x)$ is called *stable* if to every $a_1 > 0$, $b_1$, $a_2 > 0$, $b_2$ there correspond constants $a > 0$ and $b$ such that the equation

$$F(a_1x + b_1) * F(a_2x + b_2) = F(ax + b) \tag{2}$$

holds.

It is easy to verify that the normal and the improper laws are stable.

Obviously, within one type either all laws are stable or none is stable. Therefore it is possible to speak of *stable types* of laws.†

The importance of the class of stable laws for our problem is determined by the following theorem (see [73] and [59]).

Theorem. *In order that the distribution function $F(x)$ be a limit distribution for sums (1) of independent and identically distributed summands, it is necessary and sufficient that it be stable.*

*Proof.* We suppose first that the distribution functions of the sums $\zeta_n$ converge to a certain distribution function $F(x)$.

---

* I.e., such that for every $x$

$$\mathsf{P}\{\xi_1 < x\} = \mathsf{P}\{\xi_2 < x\} = \dots = \mathsf{P}\{\xi_n < x\} = \dots = F(x).$$

† The definition of a stable type can be formulated more briefly as follows: *a type is stable if it contains all the compositions of the laws belonging to it.*

According to the Lemma of § 29, if the function $F(x)$ is proper, then as $n \to \infty$

1) $$B_n \to \infty,$$

2) $$\frac{B_{n+1}}{B_n} \to 1.$$

Now let $a_1$ and $a_2$ be any two constants $(0 < a_1 < a_2 < \infty)$; then for every $\epsilon > 0$ and $n \geq n_0(\epsilon)$ it is possible to pick an index $m = m(n)$ so that

$$0 \leqslant \frac{B_m}{B_n} - \frac{a_2}{a_1} < \varepsilon.$$

Further, let $b_1$ and $b_2$ be arbitrary real constants.

Consider the sum

$$\frac{B_n}{B}\left(\frac{\xi_1+\xi_2+\cdots+\xi_n}{B_n} - A_n - b_1\right) + \frac{B_m}{B}\left(\frac{\xi_{n+1}+\cdots+\xi_{n+m}}{B_m} - A_m - b_2\right) = \frac{\xi_1+\xi_2+\cdots+\xi_{n+m}}{B} - A, \quad (3)$$

where

$$B = \frac{B_n}{a_1}, \quad A = \frac{B_nA_n + B_mA_m + b_1B_n + b_2B_m}{B}.$$

Since by hypothesis the distribution functions of the sums (1) converge to $F(x)$ as $n \to \infty$, by Theorem 2 of § 10 the distribution functions of the first and second summands on the left side of (3) converge respectively to $F(a_1^{-1}x + b_1)$ and $F(a_2^{-1}x+b_2)$. It follows that the distribution functions of the sums on the right side of (3) must converge to a limit distribution function. On the one hand, this limit must be

$$F(a_1^{-1}x + b_1) * F(a_2^{-1}x + b_2),$$

on the other hand, it must be of the same type as $F(x)$.*

In other words, we have proved that if $F(x)$ is a proper law, then it satisfies equation (2) and so is stable.

Since all improper laws are stable, the theorem is proved in one direction. The converse proposition, that every stable law $V(x)$ is the limit of distribution functions of the sums (1) of identically distributed summands, is readily deduced from the definition of a stable law. Indeed, let the variables $\xi_n$ $(k = 1, 2, \ldots)$ be independent and distributed according to the law $F(x)$; then the sum $\xi_1 + \xi_2 + \cdots + \xi_n$ is distributed according to the law $F(a_nx + b_n)$, and so the variable

$$\frac{\xi_1+\xi_2+\cdots+\xi_n}{a_n} - \frac{b_n}{a_n}$$

is distributed according to the law $F(x)$. Q.E.D.

Among the great number of results in this section we shall especially emphasize the following.

The theorem in § 34 gives an explicit form for the characteristic function of a stable distribution. Among infinitely divisible laws, and even

---

* *Translator's note.* By Theorem 1 of § 10.

among laws of the class $L$, the stable laws occupy a modest place in extent: the set of stable types depends on two real parameters $\alpha$ $(0 < \alpha \leq 2)$ and $\beta$ $(-1 \leq \beta \leq 1)$, while the set of infinitely divisible types depends on the choice of a monotone function. We remark that while the theorem just mentioned solves completely the problem of determining all stable characteristic functions, explicit expressions for the stable distribution functions are known only in a small number of cases.

In § 35 is given a complete solution of the fundamental problem in the theory of limit distributions for identically distributed summands: the necessary and sufficient conditions are indicated which must be imposed on the function $F(x)$ in order that the distributions of the sums (1) converge to a limit.

Finally, in § 37 is proved a theorem of A. Ya. Khintchine which states that if we consider the convergence of the distribution functions for the sums (1) as $n$ becomes infinite, not through all possible values but only through some subsequence of them, then the class of possible limit distributions is thereby essentially widened and turns out to coincide with the class of infinitely divisible laws.

## § 34. Canonical Representation of Stable Laws

THEOREM.* *In order that the distribution function $F(x)$ be stable, it is necessary and sufficient that the logarithm of its characteristic function be represented by the formula*

$$\log f(t) = i\gamma t - c\,|t|^{\alpha}\left\{1 + i\beta \frac{t}{|t|}\,\omega(t, \alpha)\right\}, \tag{1}$$

*where $\alpha, \beta, \gamma, c$ are constants ($\gamma$ is any real number, $-1 \leq \beta \leq 1$, $0 < \alpha \leq 2$, $c \geq 0$) and*

$$\omega(t, \alpha) = \begin{cases} \operatorname{tg} \dfrac{\pi}{2}\alpha, & \text{if} \quad \alpha \neq 1, \\ \dfrac{2}{\pi} \log |t|, & \text{if} \quad \alpha = 1. \end{cases}$$

*The functions $M(u)$ and $N(u)$ and the constant $\sigma$ in P. Lévy's formula are, correspondingly:*

| $\alpha$ | $M(u)$ | $N(u)$ | $\sigma$ | |
|---|---|---|---|---|
| $0 < \alpha < 2$ | $\dfrac{c_1}{\lvert u \rvert^{\alpha}}$ | $-\dfrac{c_2}{u^{\alpha}}$ | $0$ | $c_1 \geqslant 0,\ c_2 \geqslant 0,\ c_1 + c_2 > 0$ |
| $\alpha = 2$ | $0$ | $0$ | $\geqslant 0$ | |

* A. Ya. Khintchine and P. Lévy [62].

*Proof.* In terms of characteristic functions, equation (2) of § 33 can be rewritten as

$$\log f\left(\frac{t}{a}\right) = \log f\left(\frac{t}{a_1}\right) + \log f\left(\frac{t}{a_2}\right) + i\beta t, \tag{2}$$

where $\beta = b - b_1 - b_2$. We recall that $F(x)$ is an infinitely divisible law and, consequently,

$$\log f(t) = i\gamma t - \frac{\sigma^2 t^2}{2} + \int_{-\infty}^{0} \left\{ e^{itu} - 1 - \frac{itu}{1+u^2} \right\} dM(u) + \int_{0}^{\infty} \left\{ e^{itu} - 1 - \frac{itu}{1+u^2} \right\} dN(u).$$

Elementary calculations show that

$$\begin{aligned} i\gamma_a t - \frac{\sigma^2 t^2}{2a^2} &+ \int_{-\infty}^{0} \left\{ e^{itu} - 1 - \frac{itu}{1+u^2} \right\} dM(au) + \int_{0}^{+\infty} \left\{ e^{itu} - 1 - \frac{itu}{1+u^2} \right\} dN(au) \\ &= i\gamma_{a_1} t - \frac{\sigma^2 t^2}{2a_1^2} + \int_{-\infty}^{0} \left\{ e^{itu} - 1 - \frac{itu}{1+u^2} \right\} dM(a_1 u) + \int_{0}^{\infty} \left\{ e^{itu} - 1 - \frac{itu}{1+u^2} \right\} dN(a_1 u) \\ &+ i\gamma_{a_2} t - \frac{\sigma^2 t^2}{2a_2^2} + \int_{-\infty}^{0} \left\{ e^{itu} - 1 - \frac{itu}{1+u^2} \right\} dM(a_2 u) + \int_{0}^{\infty} \left\{ e^{itu} - 1 - \frac{itu}{1+u^2} \right\} dN(a_2 u). \end{aligned}$$

Hence, because of the uniqueness of the representation of an infinitely divisible law by Lévy's formula, we conclude that

$$\sigma^2 \left( \frac{1}{a^2} - \frac{1}{a_1^2} - \frac{1}{a_2^2} \right) = 0, \tag{3}$$

$$M(au) = M(a_1 u) + M(a_2 u) \qquad (u < 0), \tag{4}$$

$$N(au) = N(a_1 u) + N(a_2 u) \qquad (u > 0). \tag{5}$$

Now suppose that $N(u) \not\equiv 0$; then in (5) $a$ cannot vanish. Otherwise, indeed, for every $u$ the equation

$$N(a_1 u) + N(a_2 u) = N(+0)$$

would hold, which is possible only if either $N(u) \equiv 0$ or $N(u) \equiv -\infty$. The second case is impossible for the function $N(u)$.

For the equations obtained above we are going to determine all the continuous nondecreasing solutions $M(u)$ and $N(u)$ satisfying the conditions $M(-\infty) = N(+\infty) = 0$ and not identically equal to zero. For this purpose we remark that it is sufficient to confine ourselves to determining either one of these two functions, say $N(u)$. From (5) we conclude by induction that whatever the natural number $n$ and the positive numbers $a_1$, $a_2, \ldots, a_n$ may be, the following equation holds:

$$N(au) = N(a_1 u) + N(a_2 u) + \ldots + N(a_n u) \qquad (u > 0),$$

where $a$ is some positive number, depending on $a_1, a_2, \ldots, a_n$.

Hence, in particular, for $a_1 = a_2 = \cdots = a_n = 1$, we find that

$$N(au) = nN(u),$$

where $a = a(n)$. From the last equation, putting $a\,(1/n) = 1/a(n)$, we find that

$$N\left(a\left(\frac{1}{n}\right)u\right) = \frac{1}{n}N(u).$$

Finally, for an arbitrary rational number $p/q$ we find that

$$\frac{p}{q}N(u) = pN\left(a\left(\frac{1}{q}\right)u\right) = N\left(a\left(\frac{1}{q}\right)a\,(p)\,u\right) = N\left(a\left(\frac{p}{q}\right)u\right). \quad (6)$$

It is easy to see that the function $a(p/q)$ defined on the set of rational numbers is decreasing. Hence, for every $\lambda > 0$, the limits

$$a\,(\lambda - 0) = \lim_{\frac{p}{q} \to \lambda - 0} a\left(\frac{p}{q}\right),$$

$$a\,(\lambda + 0) = \lim_{\frac{p}{q} \to \lambda + 0} a\left(\frac{p}{q}\right)$$

exist. From (6) it is not difficult to deduce that for every $\lambda > 0$, $a(\lambda - 0) = a(\lambda + 0) = a(\lambda)$. Thus, for every positive $\lambda$, the function $N(u)$ satisfies the equation

$$\lambda N(u) = N(a\,(\lambda)\,u), \qquad (7)$$

where $a(\lambda) > 0$ is a decreasing continuous function of $\lambda$.

Equation (7) permits us to obtain the following important result: *the function $N(u)$ is either different from zero everywhere or identically equal to zero.*

Indeed, supposing the contrary, that for some $u_0 > 0$ the equation $N(u_0) = 0$ holds and that $N(u_1) \neq 0$ for some $u_1 > 0$; we shall reach a contradiction. First of all, from the fact that $N(u)$ is nondecreasing and $N(+\infty) = 0$, we conclude that $N(u) = 0$ for $u > u_0$ and consequently $u_1 < u_0$. Let $u_2$ $(u_1 < u_2 \leq u_0)$ be the supremum of those $u$ for which $N(u) \neq 0$. We know that the function $N(u)$ is continuous, hence $N(u_2) = 0$. According to (7),

$$\lambda N(u_1) = N\,(au_1).$$

Let $\lambda > 1$. Since $a(1) = 1$, we must have $0 < a(\lambda) < 1$. But it follows from (7) that

$$0 = \lambda N(u_2) = N(au_2).$$

Since $au_2 < u_2$, we must have $N(au_2) < 0$. We have reached a contradiction.

We shall now determine the form of the function $N(u)$, supposing that $N(u) \not\equiv 0$.

By Theorem 1 of § 30 the function $N(u)$ has derivatives (both right and left ones) for every $u$; hence we find from (7) that

$$\lambda \frac{dN(u)}{du} = a \frac{dN(au)}{d(au)}.$$

Consequently,

$$\frac{N'(u)}{N(u)} = \frac{a \dfrac{dN(au)}{d(au)}}{N(au)}.$$

We put $u = 1$ in this equation and write $\dfrac{N'(1)}{N(1)} = -\alpha$. As a result, we arrive at the following equation for the function $N(u)$: *

$$\frac{dN(a)}{N(a)} = -\alpha \frac{da}{a}.$$

Hence

$$N(a) = -c_2 a^{-\alpha}, \tag{8}$$

where $c_2$ is a constant ($c_2 > 0$).

Since $N(+\infty) = 0$, we conclude first that $\alpha$ must be a positive number. Furthermore, the integral

$$\int_0^1 u^2 \, dN(u) = c_2 \alpha \int_0^1 u^{1-\alpha} \, du$$

must converge, so $0 < \alpha < 2$.

In exactly the same way, we find that

$$M(u) = \frac{c_1}{|u|^{\alpha_1}}, \tag{9}$$

where $c_1 > 0$, $0 < \alpha_1 < 2$.

The equations (8) and (9) together with (4) and (5) yield, for $a_1 = a_2 = 1$,

$$\frac{1}{a^{\alpha}} = \frac{1}{a^{\alpha_1}} = 2. \tag{10}$$

---

* Here it is necessary to note the following fact: from (7) and the fact that $N(u) \not\equiv 0$ it is easy to deduce that when $\lambda$ varies in the interval $(0, \infty)$, $a(\lambda)$ takes all values between zero and infinity.

Hence we conclude that $\alpha = \alpha_1$. From (3), with $a_1 = a_2 = 1$, we find that

$$\sigma^2\left(\frac{1}{a^2} - 2\right) = 0.$$

If the function $M(u)$ [or $N(u)$] does not vanish, then by (10), $1/a^2 \neq 2$ and so we must have $\sigma = 0$. If, however, $\sigma \neq 0$, then $1/a^2 = 2$, and so we must have $c_1 = c_2 = 0$.

Collecting the results obtained, we see that the logarithm of the characteristic function of a stable law is either

$$\log f(t) = i\gamma t - \frac{\sigma^2}{2} t^2 \quad (\sigma > 0), \tag{11}$$

which yields the normal law, or

$$\log f(t) = i\gamma t + c_1 \int_{-\infty}^{0} \left\{ \qquad \right\} \frac{du}{|u|^{1+\alpha}}$$

$$+ c_2 \int_0^{\infty} \left\{ e^{itu} - 1 - \frac{itu}{1+u^2} \right\} \frac{du}{u^{1+\alpha}}, \tag{12}$$

where $c_1 \geq 0$, $c_2 \geq 0$, $0 < \alpha < 2$.

Elementary verification shows that the formulas (11) and (12) indeed give stable laws. Thus the problem concerning the canonical representation of stable laws is completely solved. The integrals in (12) can be expressed by elementary functions; we now proceed to the derivations of these expressions. We have to consider three cases.

1. $0 < \alpha < 1$. Since in this case the integrals

$$\int_{-\infty}^{0} \frac{u}{1+u^2} \frac{du}{|u|^{1+\alpha}} \quad \text{and} \quad \int_0^{\infty} \frac{u}{1+u^2} \frac{du}{u^{1+\alpha}}$$

are finite, (12) can be written as

$$\log f(t) = i\gamma' t + c_1 \int_{-\infty}^{0} (e^{itu} - 1) \frac{du}{|u|^{1+\alpha}} + c_2 \int_0^{\infty} (e^{itu} - 1) \frac{du}{u^{1+\alpha}}.$$

We suppose first that $t > 0$, then

$$\log f(t) = i\gamma' t + t^\alpha \left[ c_1 \int_0^{\infty} (e^{-iv} - 1) \frac{dv}{v^{1+\alpha}} + c_2 \int_0^{\infty} (e^{iv} - 1) \frac{dv}{v^{1+\alpha}} \right]. \tag{13}$$

We make use of Cauchy's theorem on contour integration, taking as the contour of integration the segment of the real axis from $O$ to $R$, the circular arc of radius $R$ with center at the origin, and the segment of the imaginary axis from $O$ to $iR$. Letting $R$ tend to infinity, we arrive at the equation

$$\int_0^\infty (e^{iv}-1)\frac{dv}{v^{1+\alpha}} = \int_0^{i\infty} (e^{iv}-1)\frac{dv}{v^{1+\alpha}} = i^{-\alpha}\int_0^\infty (e^{-y}-1)\frac{dy}{y^{1+\alpha}} = e^{-i\frac{\pi}{2}\alpha}L(\alpha),$$

where

$$L(\alpha) = \int_0^\infty (e^{-y}-1)\frac{dy}{y^{1+\alpha}} < 0.$$

Since the first integral in (13) is the conjugate complex of the second integral, it follows that

$$\int_0^\infty (e^{-iv}-1)\frac{dv}{v^{1+\alpha}} = e^{+i\frac{\pi}{2}\alpha}L(\alpha).$$

Thus, for $t > 0$,

$$\log f(t) = it\gamma' + t^\alpha L(\alpha)\left\{(c_1+c_2)\cos\frac{\pi}{2}\alpha + i(c_1-c_2)\sin\frac{\pi}{2}\alpha\right\}.$$

Noting that $\cos(\pi/2)\alpha > 0$ and putting

$$c = -L(\alpha)(c_1+c_2)\cos\frac{\pi}{2}\alpha \qquad (c \geqslant 0),$$

$$\beta = \frac{c_1-c_2}{c_1+c_2} \qquad (-1 \leqslant \beta \leqslant 1),$$

we find that

$$\log f(t) = i\gamma' t - ct^\alpha\left\{1 + i\beta\,\mathrm{tg}\,\frac{\pi}{2}\alpha\right\}$$

For $t < 0$

$$\log f(t) = \log \overline{f(-t)} = -i\gamma'(-t) - c(-t)^\alpha\left\{1 - i\beta\,\mathrm{tg}\,\frac{\pi}{2}\alpha\right\}$$
$$= i\gamma' t - c|t|^\alpha\left\{1 + i\beta\frac{t}{|t|}\,\mathrm{tg}\,\frac{\pi}{2}\alpha\right\}.$$

Therefore, for every $t$,

$$\log f(t) = i\gamma' t - c|t|^\alpha\left\{1 + i\beta\frac{t}{|t|}\,\mathrm{tg}\,\frac{\pi}{2}\alpha\right\}. \tag{14}$$

2. $\underline{1 < \alpha < 2.}$ In this case, by changing the constant $\gamma$ (12) can be written as

$$\log f(t) = it\gamma'' + c_1\int_{-\infty}^0 (e^{itu}-1-itu)\frac{du}{|u|^{1+\alpha}}$$
$$+ c_2\int_0^\infty (e^{itu}-1-itu)\frac{du}{u^{1+\alpha}}.$$

For $t > 0$,

$$\log f(t) = it\gamma'' + t^\alpha\left\{c_1\int_0^\infty (e^{-iv}-1+iv)\frac{dv}{v^{1+\alpha}} +\right.$$
$$\left. + c_2\int_0^\infty (e^{iv}-1-iv)\frac{dv}{v^{1+\alpha}}\right\}.$$

Applying to the integrals on the right side Cauchy's theorem and taking the same contour of integration as in the first case, we obtain

$$\int_0^\infty (e^{iv} - 1 - iv) \frac{dv}{v^{1+\alpha}} = i^{-\alpha} \int_0^\infty (e^{-y} - 1 + y) \frac{dy}{y^{1+\alpha}} = e^{-\frac{\pi}{2} i\alpha} M(\alpha),$$

where

$$M(\alpha) = \int_0^\infty (e^{-y} - 1 + y) \frac{dy}{y^{1+\alpha}} > 0.$$

In exactly the same way,

$$\int_0^\infty (e^{-iv} - 1 + iv) \frac{dv}{v^{1+\alpha}} = e^{\frac{\pi}{2} i\alpha} M(\alpha).$$

In this case, putting

$$c = -M(\alpha)(c_1 + c_2) \cos \frac{\pi}{2} \alpha \qquad (c > 0)$$

and attributing the same meaning to the number $\beta$ as in the first case, we find that (14) again holds for $\log f(t)$.

3. $\underline{\alpha = 1}$. Since

$$\int_0^\infty \frac{1 - \cos z}{z^2} dz = \frac{\pi}{2},$$

we find that for $t > 0$

$$\int_0^\infty \left(e^{itu} - 1 - \frac{itu}{1+u^2}\right) \frac{du}{u^2} = \int_0^\infty \frac{\cos tu - 1}{u^2} du + i \int_0^\infty \left(\sin tu - \frac{ut}{1+u^2}\right) \frac{du}{u^2}$$

$$= -\frac{\pi}{2} t + i \lim_{\varepsilon \to +0} \left[\int_\varepsilon^\infty \frac{\sin tu}{u^2} du - t \int_\varepsilon^\infty \frac{1}{u(1+u^2)} du\right]$$

$$= -\frac{\pi}{2} t + i \lim_{\varepsilon \to +0} \left[-t \int_\varepsilon^{\varepsilon t} \frac{\sin v}{v^2} dv + t\left(\int_\varepsilon^\infty \left(\frac{\sin v}{v^2} - \frac{1}{v(1+v^2)}\right) dv\right)\right].$$

But

$$\lim_{\varepsilon \to +0} \int_\varepsilon^{\varepsilon t} \frac{\sin v}{v^2} dv = \lim_{\varepsilon \to +0} \int_\varepsilon^{\varepsilon t} \frac{dv}{v} = \log t;$$

hence

$$\int_0^\infty \left(e^{itu} - 1 - \frac{itu}{1+u^2}\right) \frac{du}{u^2} = -\frac{\pi}{2} t - it \log t + i\Gamma t,$$

where

$$\Gamma = \int_0^\infty \left(\frac{\sin v}{v^2} - \frac{1}{v(1+v^2)}\right) dv.$$

Since the first integral in (12) is the conjugate complex of the second, for $t > 0$

$$\log f(t) = i\gamma' t - (c_1 + c_2)\frac{\pi}{2}t - i(c_1 - c_2)\, t \log t.$$

For $t < 0$

$$\log f(t) = \overline{\log f(-t)} = -i\gamma'(-t) - (c_1 + c_2)\frac{\pi}{2}(-t) + i(c_1 - c_2)$$
$$= (-t)\log(-t) = i\gamma' t - (c_1 + c_2)\frac{\pi}{2}|t| - i(c_1 - c_2)\, t \log |t|.$$

Putting

$$c = (c_1 + c_2)\frac{\pi}{2}, \quad \beta = \frac{c_2 - c_1}{c_1 + c_2},$$

we find that for all $t$

$$\log f(t) = i\gamma' t - c|t|\left\{1 + i\beta\frac{t}{|t|}\frac{2}{\pi}\log|t|\right\}. \tag{15}$$

The proof of the theorem is thereby completed.

We shall agree to call $\alpha$ in formula (1) the *characteristic exponent* of the stable law.

At the present time an explicit form for a stable distribution function is known only in a few cases. Thus it has long been known that the normal law ($\alpha = 2$) and the Cauchy law ($\alpha = 1$, $\beta = 0$) are stable. Recently* N. V. Smirnov proved that the stable law for which in (31) $\alpha = \frac{1}{2}$, $\beta = 1$, $\gamma = 0$, $c = 1$ has probability density equal to

$$p\left(x; \frac{1}{2}, 1, 0, 1\right) = \begin{cases} 0 & \text{for } x < 0, \\ \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2x}} x^{-\frac{3}{2}} & \text{for } x > 0, \end{cases}$$

This law belongs to the system of Pearson curves (type V).

## § 35. Domains of Attraction for Stable Laws

Let the random variables

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

be independent and have a common distribution function $F(x)$.

If for suitably chosen constants $A_n$ and $B_n$ the distribution functions of the sums

$$\zeta_n = \frac{1}{B_n}\sum_{k=1}^{n} \xi_k - A_n$$

---

* *Translator's note.* This was also obtained by P. Lévy. Sur certains processus stochastiques homogènes, *Compositio Mathematica*, **7**, 283–339 (1940); see p. 284 and p. 294.

converge as $n \to \infty$ to a distribution function $V(x)$, then we say that $F(x)$ is *attracted* to $V(x)$. The totality of distribution functions attracted to $V(x)$ is called the *domain of attraction* of the law $V(x)$. From § 33 it is clear that all stable laws, and only these, have (non-empty) domains of attraction.

Obviously, the domains of attraction of two laws belonging to the same type coincide. Hence it is possible to speak of the domain of attraction of a type.

One of the fundamental problems in the theory of stable laws should be the determination of their domains of attraction. In this section we shall give its complete solution. As an example of the forthcoming theorems we note once again the fact mentioned before: while the normal law attracts a very wide class of distribution laws, the domains of attraction of the other stable laws consist only of those distribution laws whose character recalls the character of the attracting law.

THEOREM 1.* *The distribution function $F(x)$ belongs to the domain of attraction of a (proper) normal law if and only if as $X \to \infty$*

$$\frac{X^2 \int\limits_{|x|>X} dF(x)}{\int\limits_{|x|<X} x^2\, dF(x)} \to 0. \tag{1}$$

*Proof.* We first remark that if the variance of a proper law $F(x)$ is finite, then $F(x)$ first satisfies (1) and, second, belongs to the domain of attraction of the normal law. Indeed, for such a law, as $X \to \infty$

$$X^2 \int\limits_{|x|>X} dF(x) \leqslant \int\limits_{|x|>X} x^2\, dF(x) \to 0$$

and at the same time

$$\int\limits_{|x|<X} x^2\, dF(x) \to \int x^2\, dF(x) > 0.$$

From these two relations (1) follows.

Furthermore, we put

$$a = \int x\, dF(x), \quad \sigma^2 = \int (x-a)^2\, dF(x), \quad B_n^2 = n\sigma^2.$$

Then for every $\tau > 0$ as $n \to \infty$

$$\frac{n}{B_n^2} \int\limits_{|x|>\tau B_n} (x-a)^2\, dF(x) = \frac{1}{\sigma^2} \int\limits_{|x|>\tau\sigma\sqrt{n}} (x-a)^2\, dF(x) \to 0.$$

* A. Ya. Khintchine [52], W. Feller [27], P. Lévy [75].

From Theorem 3 of § 21 it follows that $F(x)$ belongs to the domain of attraction of the normal law.

We can therefore confine ourselves to the consideration of laws $F(x)$ with infinite variances. In other words, we can confine ourselves to the consideration of laws $F(x)$ for which, as $X \to \infty$,

$$\int_{-X}^{X} x^2\, dF(x) \to \infty. \tag{2}$$

We shall prove first that if (2) holds, then as $X \to \infty$

$$\left[\int_{-X}^{X} x\, dF(x)\right]^2 = o\left[\int_{-X}^{X} x^2\, dF(x)\right]. \tag{3}$$

Indeed, let $z(x)$ be a positive function which is unbounded as $x \to \pm\infty$ and such that the integral

$$C = \int z^2(x)\, dF(x)$$

is finite. Then by the inequality of Cauchy and Bunyakovskii

$$\left[\int_{-X}^{X} x\, dF(x)\right]^2 = \left[\int_{-X}^{X} z(x) \frac{x}{z(x)}\, dF(x)\right]$$
$$\leqslant \int_{-X}^{X} z^2(x)\, dF(x) \int_{-X}^{X} \frac{x^2}{z^2(x)}\, dF(x) \leqslant C \int_{-X}^{X} \frac{x^2}{z^2(x)}\, dF(x).$$

From this inequality (3) obviously follows.

According to Theorem 4 of § 26 the law $F(x)$ belongs to the domain of attraction of the normal law if and only if there exists a sequence of constants $C_n$ ($C_n \to \infty$ as $n \to \infty$) such that as $n \to \infty$

$$n \int_{|x| > C_n} dF(x) \to 0, \tag{4}$$

$$\frac{n}{C_n^2}\left\{\int_{|x| < C_n} x^2\, dF(x) - \left(\int_{|x| < C_n} x\, dF(x)\right)^2\right\} \to \infty.$$

By what has just been proved for laws with infinite variances the second condition can be replaced by a simpler one, namely, as $n \to \infty$

$$\frac{n}{C_n^2} \int_{|x| < C_n} x^2\, dF(x) \to \infty. \tag{5}$$

We shall first prove that (4) and (5) imply (1). To this end we note that since $C_n \to \infty$ as $n \to \infty$, for every sufficiently large $X$ an $n$ can be found such that

$$C_n \leqslant X < C_{n+1}.$$

For the sake of brevity we introduce the notations

$$\chi(X) = \int\limits_{|x|>X} dF(x), \quad H(X) = \frac{1}{X^2} \int\limits_{|x|<X} x^2 \, dF(x).$$

It is easy to see that

$$H(C_n) + \chi(C_n) \geqslant H(X) \geqslant H(C_{n+1}) - \chi(C_n),$$

and also

$$\chi(C_n) \geqslant \chi(X) \geqslant \chi(C_{n+1}).$$

From these inequalities we find that

$$\frac{n\chi(C_n)}{\frac{n}{n+1}(n+1)H(C_{n+1}) - n\chi(C_n)} \geqslant \frac{\chi(X)}{H(X)} \geqslant \frac{(n+1)\chi(C_{n+1})}{\frac{n+1}{n}[nH(C_n) + n\chi(C_n)]}.$$

By (4) and (5) the first and last fractions tend to zero as $n \to \infty$ and so (1) holds.

We shall now prove that the condition (1) is sufficient, i.e., it is possible to determine a sequence of constants $C_n$ ($C_n \to \infty$ as $n \to \infty$) for which (4) and (5) will be satisfied simultaneously. To this end we pick an arbitrary $\delta > 0$ and denote by $C_n(\delta)$ the infimum of all $X$ for which

$$n\chi(X) \leqslant \delta. \tag{6}$$

Since by hypothesis

$$\int x^2 \, dF(x) = +\infty,$$

it is evident that $C_n(\delta) \to \infty$ as $n \to \infty$. By (1), whatever $\epsilon$ may be, for $n \geq n(\delta,\epsilon)$

$$nH\left(\frac{1}{2} C_n(\delta)\right) > \frac{\delta}{\varepsilon}.$$

Since

$$H\left(\frac{1}{2} C_n(\delta)\right) = \frac{4}{C_n^2(\delta)} \int\limits_{|x| < \frac{C_n(\delta)}{2}} x^2 \, dF(x)$$

$$\leqslant \frac{4}{C_n^2(\delta)} \int\limits_{|x| < C_n(\delta)} x^2 \, dF(x) = 4H(C_n(\delta))$$

for $n \geq n(\delta,\epsilon)$,

$$nH(C_n(\delta)) > \frac{\delta}{4\varepsilon}.$$

Therefore, for every $\delta > 0$, as $n \to \infty$

$$nH(C_n(\delta)) \to \infty.$$

Consequently, it is possible to pick a sequence $\delta_n$ converging to zero so that

$$nH(C_n(\delta_n)) \to \infty \qquad \text{as} \quad n \to \infty.$$

But by (6),

$$n\chi(C_n(\delta_n)) \leqslant \delta_n \to 0 \qquad \text{as} \quad n \to \infty.$$

In other words, as $n \to \infty$

$$n \int_{|x| > C_n} dF(x) \to 0$$

and

$$\frac{n}{C_n^2} \int_{|x| < C_n} x^2\, dF(x) \to \infty.$$

We have proved that (1) implies (4) and (5); the theorem is proved.

THEOREM 2.* *In order that the distribution function $F(x)$ belong to the domain of attraction of a stable law with the characteristic exponent $\alpha$ $(0 < \alpha < 2)$* † *it is necessary and sufficient that*

1) $\dfrac{F(-x)}{1-F(x)} \to \dfrac{c_1}{c_2}$ as $x \to \infty$. (7)

2) *for every constant $k > 0$*

$$\frac{1 - F(x) + F(-x)}{1 - F(kx) + F(-kx)} \to k^{\alpha} \qquad \text{as} \quad x \to \infty. \tag{8}$$

*Proof.* According to the Theorem 4 of § 25, in order that the law $F(x)$ belong to the domain of attraction of a stable law with exponent $\alpha$ $(0 < \alpha < 2)$, it is necessary and sufficient that for some choice of the constants $B_n$ the following conditions be satisfied:

---

* B. V. Gnedenko [38], Doeblin [24].

† *Translator's note.* It should be noted that this theorem also gives a necessary and sufficient condition for $F(x)$ to belong to the domain of attraction of any given stable law determined by the constants $\alpha$, $c_1$, and $c_2$ subject to the necessary conditions: $0 < \alpha < 2$, $c_1 + c_2 > 0$, $|c_1 - c_2| \leq c_1 + c_2$ (see § 34). Also, the proper choice of the normalizing constants can be deduced from the proof. For $B_n$ this is stated explicitly. For $A_n$ it can be shown (see Theorem 4 of § 25) that it may be chosen to be $\dfrac{n}{B_n}\displaystyle\int_{-\infty}^{\infty} x\, dF(x)$ if $1 < \alpha \leq 2$; $n \operatorname{Im} \log f(1/B_n)$ if $\alpha = 1$, where $f$ is the characteristic function of $F$; and $= 0$ if $\alpha < 1$. Cf. B. V. Gnedenko and V. S. Korolyuk, Some remarks on the theory of domains of attraction of stable distributions (in Russian), *Doklady Akad. Nauk Ukrain. SSR*, no. 4, 275–278 (1950), where the conditions 1) and 2) in Theorems are also given in terms of the characteristic function.

$$nF(B_n x) \to \frac{c_1}{|x|^\alpha} \qquad (x < 0), \tag{9}$$

$$n\Big(1 - F(B_n x)\Big) \to \frac{c_2}{x^\alpha} \qquad (x > 0), \tag{10}$$

$$\lim_{\varepsilon \to 0} \overline{\lim_{n \to \infty}}\, n \Big\{ \int\limits_{|x|<\varepsilon} x^2 \, dF(B_n x) - \Big( \int\limits_{|x|<\varepsilon} x \, dF(B_n x) \Big)^2 \Big\} = 0. \tag{11}$$

The *necessity* of the conditions of the theorem follows from these relations without difficulty. In fact, let $y > 0$ be large. Choose $n$ so that for a given $x > 0$

$$B_n x \leqslant y \leqslant B_{n+1} x.$$

Obviously,

$$F(-B_{n+1}x) \leqslant F(-y) \leqslant F(-B_n x),$$
$$1 - F(B_{n+1}x) \leqslant 1 - F(y) \leqslant 1 - F(B_n x),$$

and for $k > 0$

$$F(-kB_{n+1}x) \leqslant F(-ky) \leqslant F(-B_n kx),$$
$$1 - F(B_{n+1}kx) \leqslant 1 - F(ky) \leqslant 1 - F(B_n kx).$$

Hence

$$\frac{F(-B_{n+1}x)}{1 - F(B_n x)} \leqslant \frac{F(-y)}{1 - F(y)} \leqslant \frac{F(-B_n x)}{1 - F(B_{n+1}x)},$$

and also

$$\frac{1 - F(B_{n+1}x) + F(-B_{n+1}x)}{1 - F(B_n kx) + F(-B_n kx)} \leqslant \frac{1 - F(y) + F(-y)}{1 - F(ky) + F(-ky)}$$

$$\leqslant \frac{1 - F(B_n x) + F(-B_n x)}{1 - F(B_{n+1}kx) + F(-B_{n+1}kx)}.$$

An application of (9) and (10) gives

$$\frac{c_1}{c_2} \leqslant \lim_{y \to \infty} \frac{F(-y)}{1 - F(y)} \leqslant \frac{c_1}{c_2},$$

$$k^\alpha \leqslant \lim_{y \to \infty} \frac{1 - F(y) + F(-y)}{1 - F(ky) + F(-ky)} \leqslant k^\alpha.$$

*Sufficiency.* The conditions of the theorem have meaning only if for every $x$

$$1 - F(x) + F(-x) > 0,$$

i.e.,

$$\mathsf{P}\{|\xi| > x\} > 0.$$

Hence it follows that the infimum of all $x$ satisfying the inequality

$$\mathsf{P}\{|\xi| > x\} \leqslant \frac{c_1 + c_2}{n} \leqslant \mathsf{P}\{|\xi| \geqslant x\}, \tag{12}$$

which we denote by $B_n$, tends to infinity as $n \to \infty$.

From (8) it follows that * for every $x > 0$

$$n(1 - F(xB_n) + F(-xB_n)) = n(1 - F(B_n + 0) + F(-B_n - 0))\frac{1 + o(1)}{x^\alpha},$$

$$n(1 - F(xB_n) + F(-xB_n)) = n(1 - F(B_n - 0) + F(-B_n + 0))\frac{1 + o(1)}{x^\alpha}.$$

From this and (12) we conclude that as $n \to \infty$

$$n(1 - F(xB_n) + F(-xB_n)) \to \frac{c_1 + c_2}{x^\alpha}. \tag{13}$$

From (7) it follows that for every $x > 0$

$$c_1 n(1 - F(xB_n)) = c_2 nF(-xB_n)(1 + o(1)). \tag{14}$$

From (13) and (14) follow (9) and (10). It remains to derive (11) from (7) and (8).

For this purpose we first establish that

$$\int x^2\, dF(x) = +\infty. \tag{15}$$

Let $x_0$ be so large that for a given $\epsilon > 0$ and $k > 1$, chosen so that $k^{2-\alpha}(1 - \epsilon) > 1$ and $k^\alpha(1 - \epsilon) > 1$,

$$\frac{\mathsf{P}\{|\xi| > k^s x_0\}}{\mathsf{P}\{|\xi| > k^{s+1}x_0\}} = (1 + \varepsilon_s)^{-1} k^\alpha, \tag{16}$$

where $|\epsilon_s| \leq \epsilon$ for $s = 0, 1, 2, \dots$.

By the condition 2) of the theorem such an $x_0$ can be found. Obviously,

$$\int x^2\, dF(x) = \int\limits_{|x| < x_0} x^2\, dF(x) + \sum_{s=1}^{\infty} \int\limits_{k^{s-1}x_0 \leqslant |x| < k^s x_0} x^2\, dF(x)$$

$$\geqslant \int\limits_{|x| < x_0} x^2\, dF(x) + x_0^2 \sum_{s=1}^{\infty} k^{2(s-1)} \mathsf{P}\{k^{s-1}x_0 \leqslant |\xi| < k^s x_0\}.$$

By (16)

$$\mathsf{P}\{k^{s-1}x_0 \leqslant |\xi| < k^s x_0\} = \mathsf{P}\{|\xi| > k^{s-1}x_0\} - \mathsf{P}\{|\xi| > k^s x_0\}$$

$$\geqslant \mathsf{P}\{|\xi| > x_0\}\left\{k^{-\alpha(s-1)} \prod_{r=1}^{s-1}(1 + \varepsilon_r) - k^{-\alpha s}\prod_{r=1}^{s}(1 + \varepsilon_r)\right\}$$

$$\geqslant \mathsf{P}\{|\xi| > x_0\}(1 - \varepsilon)^s k^{-\alpha s}\left\{\frac{k^\alpha}{1 + \varepsilon} - 1\right\}.$$

Thus,

$$\int x^2\, dF(x) \geqslant \int\limits_{|x| < x_0} x^2\, dF(x) + x_0^2 \mathsf{P}\{|\xi| > x_0\} k^{-2}\left\{\frac{k^\alpha}{1 + \epsilon} - 1\right\}\sum_{s=1}^{\infty}(1 - \varepsilon)^s k^{s(2-\alpha)}.$$

---

* *Translator's note.* Apply (8) with, e.g., $x\left(B_n \pm \frac{1}{n}\right)$ for $x$ and $\frac{1}{x}$ for $k$.

Since $\mathbf{P}\{|\xi| > x_0\} > 0$ for every $x_0$ and the series on the right side of the inequality diverges, (15) has been proved.

Let $x_0$ be chosen as before. Choose $n$ so large that

$$\int_{|x|<x_0} x^2\,dF(x) \leqslant \int_{x_0 \leqslant |x| < B_n\varepsilon} x^2\,dF(x);$$

this is possible by (15); then obviously

$$\int_{|x|<B_n\varepsilon} x^2\,dF(x) \leqslant 2 \int_{x_0 \leqslant |x| < B_n\varepsilon} x^2\,dF(x).$$

Let the integer $s > 0$ be such that

$$k^s x_0 \leqslant B_n\varepsilon < k^{s+1} x_0. \tag{17}$$

We have

$$\int_{|x|<B_n\varepsilon} x^2\,dF(x) \leqslant 2x_0^2 \sum_{r=0}^{s} k^{2(r+1)} \mathbf{P}\{k^r x_0 \leqslant |\xi| < k^{r+1} x_0\}$$
$$< 2x_0^2 \sum_{r=0}^{s} k^{2(r+1)} \mathbf{P}\{|\xi| \geqslant k^r x_0\} \leqslant 2\varepsilon^2 B_n^2 \sum_{r=0}^{s} k^{2(r-s+1)} \mathbf{P}\{|\xi| \geqslant k^r x_0\}. \tag{18}$$

By (16) and (17),

$$\mathbf{P}\{|\xi| \geqslant k^r x_0\} \leqslant k^\alpha (1+\varepsilon) \mathbf{P}\{|\xi| \geqslant k^{r+1} x_0\}$$
$$\leqslant [(1+\varepsilon)\,k^\alpha]^{s-r+1} \mathbf{P}\{|\xi| \geqslant k^{s+1} x_0\} \leqslant [(1+\varepsilon)\,k^\alpha]^{s-r+1} \mathbf{P}\{|\xi| \geqslant \varepsilon B_n\}.$$

From (18) we find that for $k^{\alpha-2} < (1+\epsilon)^{-1}$

$$\int_{|x|<\varepsilon B_n} x^2\,dF(x)$$
$$\leqslant 2B_n^2 \varepsilon^2 (1+\varepsilon)\,k^{\alpha+2} \mathbf{P}\{|\xi| \geqslant \varepsilon B_n\} \sum_{r=0}^{s} [(1+\varepsilon)\,k^{\alpha-2}]^{s-r}$$
$$\leqslant 2B_n^2 \varepsilon^2 (1+\varepsilon)\,k^{\alpha+2} \mathbf{P}\{|\xi| \geqslant \varepsilon B_n\} \frac{1}{1-(1+\varepsilon)\,k^{\alpha-2}}.$$

Hence by (13),

$$\overline{\lim_{n\to\infty}} \frac{n}{B_n^2} \int_{|x|<\varepsilon B_n} x^2\,dF(x) \leqslant 2\varepsilon^{2-\alpha}(c_1+c_2) \frac{(1+\varepsilon)\,k^{\alpha+2}}{1-(1+\varepsilon)\,k^{\alpha-2}}$$

and therefore

$$\lim_{\varepsilon\to 0} \overline{\lim_{n\to\infty}} \frac{n}{B_n^2} \int_{|x|<\varepsilon B_n} x^2\,dF(x) = 0.$$

Now the inequality

$$\frac{n}{B_n^2}\left\{\int_{|x|<\epsilon B_n} x^2\,dF(x) - \left(\int_{|x|<\epsilon B_n} x\,dF(x)\right)^2\right\} \leqslant \frac{n}{B_n^2} \int_{|x|<\epsilon B_n} x^2\,dF(x)$$

proves the validity of the asserted theorem.

THEOREM 3.* *If the distribution function $F(x)$ belongs to the domain of attraction of a stable law with exponent $\alpha$, then for every $\delta$ $(0 \leq \delta < \alpha)$ the integral*

$$\int |x|^{\delta} dF(x)$$

*exists.*

*Proof.* We consider first the case $\alpha < 2$. Then, whatever $\epsilon > 0$ and $k > 1$ may be, according to the preceding theorem it is possible to find an $x_0$ such that

$$\frac{\mathsf{P}\{|\xi| > k^s x_0\}}{\mathsf{P}\{|\xi| > k^{s+1} x_0\}} = (1 + \varepsilon_s)^{-1} k^{\alpha}, \tag{19}$$

where $|\epsilon_s| \leq \epsilon$ for $s = 0, 1, 2, \ldots$. We suppose that $\frac{1+\epsilon}{k^{\alpha-\delta}} < 1$. Obviously,

$$\int |x|^{\delta} dF(x) \leqslant{}^{\dagger} \int_{-x_0}^{x_0} |x|^{\delta} dF(x) + \sum_{s=1}^{\infty} \int\limits_{k^{s-1}x_0 \leqslant |x| < k^s x_0} |x|^{\delta} dF(x)$$

$$\leqslant \int_{-x_0}^{x_0} |x|^{\delta} dF(x) + x_0^{\delta} \sum_{s=1}^{\infty} k^{s\delta} \mathsf{P}\{|\xi| > k^{s-1} x_0\}. \tag{20}$$

By (19),

$$\mathsf{P}\{|\xi| > k^s x_0\} = \prod_{r=0}^{s-1} (1 + \varepsilon_r) \cdot k^{-\alpha s} \mathsf{P}\{|\xi| > x_0\},$$

hence

$$\sum_{s=1}^{\infty} k^{s\delta} \mathsf{P}\{|\xi| > k^{s-1} x_0\} \leqslant \mathsf{P}\{|\xi| > x_0\} k^{\alpha} \sum_{s=1}^{\infty} (1 + \varepsilon)^s k^{-s(\alpha-\delta)}$$

and so the right side of (20) is finite by the choice of $\epsilon$ and $k$.

For the proof of the theorem in the case $\alpha = 2$ we consider the function

$$\psi(z) = \int\limits_{|x| \leqslant z} x^2 dF(x) = -z^2 \int\limits_{|x| > z} dF(x) + 2 \int_0^z v\, dv \int\limits_{|x| > v} dF(x).$$

---

* For the case $\alpha = 2$ see, e.g., A. Ya. Khintchine [52], H. Cramér [21]. For the case $\alpha < 2$ see B. V. Gnedenko [44].

† *Translator's note.* In the original an equality sign stands here. However, according to the convention used in this book (Chapter 1, § 6) $\int_{-x_0}^{x_0} |x|^{\delta}\, dF(x)$ stands for $\int\limits_{-x_0 \leq x < x_0} |x|^{\delta}\, dF(x)$. Hence an inequality sign is in order. We have not undertaken to check all similar formulas.

Obviously it is sufficient to confine our consideration to the case that the variance of $F(x)$ is infinite.

In this case the nondecreasing function $\psi(z)$ becomes infinite as $z \to \infty$. By Theorem 1,

$$\psi(z) \leqslant 2 \int_0^z v\,dv \int_{|x|>v} dF(x) = o\left(\int_1^z \frac{\psi(v)}{v}\,dv\right).$$

Let $M(z)$ denote the supremum of $v^{-\epsilon}\psi(v)$ in the interval $1 \leq v \leq z$, where $\epsilon > 0$ is a given number. Then

$$\int_1^z \frac{\psi(v)}{v}\,dv \leqslant M(z) \int_1^z v^{\varepsilon-1}dv < \frac{z^{\varepsilon} M(z)}{\varepsilon}$$

and consequently,

$$z^{-\varepsilon}\psi(z) = o(M(z)).$$

From this we can obviously conclude that $\psi(z) = o(z^{\epsilon})$ for every $\epsilon > 0$. But for $0 < \delta < 2$ and all sufficiently large $z > 0$

$$\int_{z \leqslant |x| < 2z} |x|^{\delta} dF(x) < z^{-(2-\delta)}\psi(2z) < z^{\frac{\delta}{2}-1},$$

so that

$$\sum_{k=0}^{\infty} \int_{2^k z \leqslant |x| \leqslant 2^{k+1} z} |x|^{\delta} dF(x) < z^{\frac{\delta}{2}-1} \sum_{k=0}^{\infty} 2^{k\left(\frac{\delta}{2}-1\right)} < +\infty.$$

The theorem is thereby completely proved.

If the normalizing coefficients $B_n$ in the sums

$$\zeta_n = \frac{\xi_1 + \xi_2 + \dots + \xi_n}{B_n} - A_n$$

are not chosen arbitrarily, but are subject to some restriction on their rate of increase, then it is clear that the domain of attraction under the given restriction can only be narrower than without this restriction. For example, the theorem of de Moivre-Laplace shows that in the investigation of the domain of attraction of the normal law the following choice of the normalizing coefficients is of special interest:

$$B_n = a\sqrt{n},$$

where $a$ is a constant.

If $V(x)$ is any given stable law, then it is easy to see that it belongs to its own domain of attraction. If $\alpha$ is the characteristic exponent of the law $V(x)$, then the normalizing coefficients are determined by the formula

$$B_n = n^{\frac{1}{\alpha}}.$$

*We say that the law $F(x)$ belongs to the domain of normal attraction of the law $V(x)$, if for some $a > 0$ and some $A_n$ the following relation holds:*

$$\lim_{n\to\infty} \mathsf{P}\left\{\frac{1}{an^{\frac{1}{\alpha}}}\sum_{k=1}^{n}\xi_k - A_n < x\right\} = V(x). \tag{21}$$

Here $\alpha$ is the characteristic exponent of the law $V(x)$.

THEOREM 4. *In order that the law $F(x)$ belong to the domain of normal attraction of the law $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{z^2}{2}}\,dz$, it is necessary and sufficient that it have a finite variance. If this be so, then necessarily $a = \sigma$ in (21), and $A_n$ may be chosen to be $n\int x\,dF(x)$.*

*Proof*. According to Theorem 4 of § 26, for attraction to the law $\Phi(x)$ it is necessary and sufficient that for every $\epsilon > 0$, as $n \to \infty$,

$$\left.\begin{aligned} &n\int_{|x|>\epsilon B_n} dF(x) \to 0,\\ &\frac{n}{B_n^2}\left[\int_{|x|<\epsilon B_n} x^2\,dF(x) - \left(\int_{|x|<\epsilon B_n} x\,dF(x)\right)^2\right] \to 1. \end{aligned}\right\} \tag{22}$$

In case of normal attraction, the second of these conditions becomes

$$\lim_{n\to\infty}\left[\int_{|x|<\epsilon a\sqrt{n}} x^2\,dF(x) - \left(\int_{|x|<\epsilon a\sqrt{n}} x\,dF(x)\right)^2\right] = \sigma^2.$$

By Theorem 3, the limit

$$\lim_{n\to\infty}\int_{|x|<\epsilon a\sqrt{n}} x\,dF(x) = \int x\,dF(x)$$

exists; hence

$$\int\left(x - \int x\,dF(x)\right)^2 dF(x) = \sigma^2 \tag{23}$$

also exists. The converse proposition, that the finiteness of the integral (23) implies (22), is trivial.

THEOREM 5.* *In order that the law $F(x)$ belong to the domain of normal attraction of the stable law $V(x)$ with characteristic exponent $\alpha$ $(0 < \alpha < 2)$ and given constants $c_1$ and $c_2$, it is necessary and sufficient that*

---

* B. V. Gnedenko [38].

$$\left.\begin{aligned} F(x) &= (c_1a^\alpha + \alpha_1(x))\frac{1}{|x|^\alpha} && \text{for} \quad x < 0, \\ F(x) &= 1 - (c_2a^\alpha + \alpha_2(x))\frac{1}{x^\alpha} && \text{for} \quad x > 0, \end{aligned}\right\} \tag{24}$$

*where a is a positive constant and the functions $\alpha_1(x)$ and $\alpha_2(x)$ satisfy the conditions*

$$\lim_{x \to -\infty} \alpha_1(x) = \lim_{x \to \infty} \alpha_2(x) = 0. \tag{25}$$

*The constant a in* (21) *and* (24) *is the same.*

*Proof. Necessity.* According to Theorem 4 of § 25, necessary and sufficient conditions for the attraction of the law $F(x)$ to the stable law determined by the constants $c_1 \geq 0$, $c \geq 0$, and $\alpha$ in formula (12) of § 34, are that as $n \to \infty$

$$\left.\begin{aligned} nF(B_nx) &\to \frac{c_1}{|x|^\alpha} && \text{for} \quad x < 0, \\ n(1 - F(B_nx)) &\to \frac{c_2}{x^\alpha} && \text{for} \quad x > 0. \end{aligned}\right\} \tag{26}$$

Since for normal attraction $B_n = an^{\frac{1}{\alpha}}$, by putting $y = B_nx$ we can write (24) as

$$\begin{aligned} F(y) &= (c_1a^\alpha + \alpha_1(y))\frac{1}{|y|^\alpha} && \text{for } y < 0, \\ F(y) &= 1 - (c_2a^\alpha + \alpha_2(y))\frac{1}{y^\alpha} && \text{for } y > 0, \end{aligned}$$

where $\alpha_1(y)$ and $\alpha_2(y)$ are certain functions satisfying (25). The *sufficiency* of the theorem is trivial.

## § 36. Properties of Stable Laws

The preceding results enable us to record a series of essential properties of the stable laws.

1. (P. Lévy [76], p. 201.) For every stable law $V(x)$ except the normal and the unitary, there exist numbers $\alpha$ $(0 < \alpha < 2)$ and $c > 0$ such that

$$\lim_{x \to +\infty} x^\alpha \{1 - V(x) + V(-x)\} = c. \tag{1}$$

Obviously, (1) holds for every law belonging to the domain of normal attraction of a stable law with characteristic exponent $\alpha$ $(0 < \alpha < 2)$. Since each stable law belongs to its own domain of normal attraction, it also satisfies (1).

2. (B. V. Gnedenko [38].) Every stable law with characteristic exponent $\alpha$ $(0 < \alpha < 2)$ has finite absolute moments of order $\delta$ $(0 < \delta < \alpha)$.*

---

* *Translator's note.* It should be added that, on the other hand, all absolute moments of order $\geq \alpha$ are infinite. This follows from (1) and is needed below.

This is also a consequence of the fact that every stable law belongs to its own domain of normal attraction.

Hence,* in particular, it follows that among all stable laws only the normal law has a finite variance. For $1 < \alpha < 2$ the stable laws have mathematical expectations, for $0 < \alpha \leq 1$ the stable laws have neither variance nor mathematical expectation.

3. (A. Ya. Khintchine [59], p. 101.) All proper stable laws are continuous and have derivatives of all orders at every point.

Indeed, since for a proper stable law $V(x)$

$$|v(t)| = e^{-c|t|^{\alpha}} \qquad (c > 0, \quad 0 < \alpha \leqslant 2), \tag{2}$$

the inversion formula in our case can be written as

$$V(x) - V(x_0) = \frac{1}{2\pi} \int \frac{e^{-itx_0} - e^{-itx}}{it} v(t)\, dt.$$

Differentiating this formula formally $n$ times, we find that

$$V^{(n)}(x) = \frac{(-i)^{n-1}}{2\pi} \int t^{n-1} e^{-itx} v(t)\, dt. \tag{3}$$

The integral (3) converges absolutely, thus proving our assertion.

The last theorem can be improved.

THEOREM. (A. I. Lapin [71].) *A proper stable distribution function with exponent $\alpha \geq 1$ is analytic on the entire real axis. For $\alpha > 1$ it is an entire function. For $\alpha = 1$ the radius of convergence of its Taylor series in the neighborhood of any point is not less than $c$.*

*Proof.* From (2) and (3) we deduce that for every real $x$

$$a_n = \left| \frac{V^{(n)}(x)}{n!} \right| \leqslant \frac{1}{\pi\alpha} \cdot \frac{\Gamma\left(\frac{n}{\alpha}\right)}{n!} c^{-\frac{n}{\alpha}}.$$

Hence

$$R = \frac{1}{\overline{\lim} \sqrt[n]{a_n}} = \begin{cases} \infty & \text{for} \quad \alpha > 1, \\ c & \text{for} \quad \alpha = 1. \end{cases}$$

## § 37. DOMAINS OF PARTIAL ATTRACTION

In § 33 we proved that the stable distributions and only the stable distributions can appear as the limit distributions for normalized sums

$$\zeta_n = \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{B_n} - A_n$$

of independent and identically distributed random variables, as the number $n$ of summands tends to infinity, running through all integral values. It may happen that the distribution functions of the sums $\zeta_n$ do

---

* *Translator's note.* See the preceding note.

not converge for any choice of the constants $B_n$ and $A_n$, but that for some subsequence $n_1 < n_2 < \cdots < n_k < \cdots$ there is convergence. The general theory permits us only to assert that this limit law is necessarily infinitely divisible. As A. Ya. Khintchine [58] proved, the incomparably deeper converse proposition is also true: every infinitely divisible distribution can appear as the limit for distributions of the sums $\zeta_{n_k}$.

We shall say that $F(x)$ belongs to the domain of *partial attraction of the law* $V(x)$ or, what is the same thing, of its type, if there exists a subsequence $n_1 < n_2 < \cdots < n_k < \cdots$ such that the distribution functions of the sums $\zeta_{n_k}$ for suitably chosen constants $B_n$ and $A_n$ converge to $V(x)$. In these terms we can state the preceding theorem as follows:

THEOREM. *Every infinitely divisible law has a (non-empty) domain of partial attraction.*

*Proof.* Consider the infinitely divisible law $V(x)$ for which

$$\log v(t) = i\gamma t + \int \left\{ e^{itu} - 1 - \frac{itu}{1+u^2} \right\} \frac{1+u^2}{u^2}\, dG(u).$$

Since for the normal law the theorem becomes trivial, we may suppose that the function $G(u)$ has a point of increase $u_0$ different from 0. Let $a > 0$ be such that $a^{-1} < |u_0| < a$. Consider the domain $\Delta_k$ defined by the inequalities

$$a^{-k} < |u| < a^k \qquad (k = 1, 2, \ldots),$$

and put

$$\Gamma_k = \int_{\Delta_k} \frac{1+u^2}{u^2}\, dG(u).$$

With increasing $k$ the domains $\Delta_k$ are enlarged; hence

$$0 < \Gamma_1 \leqslant \Gamma_2 \leqslant \cdots \leqslant \Gamma_k \leqslant \cdots$$

Furthermore, put

$$A_k = \int_{\Delta_k} \frac{1+u^2}{u}\, dG(u), \qquad B_k = \int_{\Delta_k} (1+u^2)\, dG(u),$$

$$C_k = \int_{\Delta_k} u(1+u^2)\, dG(u),$$

$$\sigma^2 = G(+0) - G(-0),$$

and introduce the numbers $\lambda_k$ defined as follows:

$$\lambda_1 = 1, \qquad \lambda_k = \frac{1}{\sigma^2 + \frac{1}{k}} \sum_{r=1}^{k-1} \lambda_r B_r \qquad (k > 1) \tag{1}$$

Now let the sequence of natural numbers

$$q_1 < q_2 < \dots < q_n < \dots$$

increase so fast that

$$\lim_{n\to\infty} q_n \sum_{k=n+1}^{\infty} \frac{\Gamma_k}{q_k} = 0, \tag{2}$$

$$\lim_{n\to\infty} q_n^{-\frac{1}{2}} \lambda_n^{-\frac{3}{2}} \sum_{k=1}^{n-1} q_k^{\frac{1}{2}} \lambda_k^{\frac{3}{2}} | C_k | = 0. \tag{3}$$

Also, put

$$\beta_k^2 = \lambda_k q_k .$$

We shall now turn to the determination of a law partially attracted to the law $V(x)$. To this end consider the infinitely divisible laws $\Psi_k(x)$ for which

$$\log \psi_k (t) = \int\limits_{\Delta_k} (e^{itu} - 1) \frac{1+u^2}{u^2} dG(u).$$

Since according to (2) the series $\sum\limits_k \frac{\Gamma_k}{q_k}$ converges,

$$\log f(t) = \sum_{k=1}^{\infty} \frac{1}{q_k} \log \psi_k (t\beta_k)$$

is the logarithm of some characteristic function. Putting

$$R_n = \sum_{k=n+1}^{\infty} \frac{\Gamma_k}{q_k},$$

we can obviously write

$$\log f\left(\frac{t}{\beta_n}\right) = \sum_{k=1}^{n-1} \frac{1}{q_k} \log \psi_k\left(\frac{\beta_k}{\beta_n} t\right) + \frac{1}{q_n} \log \psi_n (t) + O(R_n). \tag{4}$$

But

$$\sum_{k=1}^{n-1} \frac{1}{q_k} \log \psi_k\left(\frac{\beta_k}{\beta_n} t\right) = \sum_{k=1}^{n-1} \frac{1}{q_k} \int\limits_{\Delta_k} \left(e^{i\frac{\beta_k}{\beta_n}tu} - 1\right) \frac{1+u^2}{u^2} dG(u)$$

$$= \frac{it}{\beta_n} v_{n-1} - \frac{t^2}{2\beta_n^2} \sum_{k=1}^{n-1} \frac{\beta_k^2 B_k}{q_k} + \frac{\theta}{6} \frac{t^3}{\beta_n^3} u_{n-1}, \qquad |\theta| \leqslant 1,$$

where we have put

$$v_{n-1} = \sum_{k=1}^{n-1} \frac{\beta_k A_k}{q_k}, \quad u_{n-1} = \sum_{k=1}^{n-1} \frac{\beta_k^3 | C_k |}{q_k} = \sum_{k=1}^{n-1} q_k^{\frac{1}{2}} \lambda_k^{\frac{3}{2}} | C_k |.$$

From (4) we now find

$$\log f\left(\frac{t}{\beta_n}\right) = it\frac{v_{n-1}}{\beta_n} - \frac{t^2}{2\lambda_n q_n}\sum_{k=1}^{n-1}\lambda_k B_k + \frac{1}{q_n}\log\psi_n(t)$$

$$+ \theta t^3 q_n^{-\frac{3}{2}}\lambda_n^{-\frac{3}{2}}\sum_{k=1}^{n-1} q_k^{\frac{1}{2}}\lambda_k^{\frac{3}{2}}|C_k| + O(R_n), \qquad |\theta| \leqslant 1.$$

Using (1), (2), (3) we find that for every $t$

$$q_n \log f\left(\frac{t}{\beta_n}\right) = it\,\frac{q_n v_{n-1}}{\beta_n} - \left(\sigma^2 + \frac{1}{n}\right)\frac{t^2}{2} + \log\psi_n(t)$$

$$+ O(q_n R_n) + O\left(q_n^{-\frac{1}{2}}\lambda_n^{-\frac{3}{2}} u_{n-1}\right)$$

$$= it\frac{q_n v_{n-1}}{\beta_n} - \sigma^2\frac{t^2}{2} + \int_{\Delta_n}(e^{itu}-1)\frac{1+u^2}{u^2}\,dG(u) + o(1).$$

Writing

$$\gamma_n = \frac{q_n v_{n-1}}{\beta_n} - \gamma + \int_{\Delta_n}\frac{dG(u)}{u},$$

we find that for every $t$

$$q_n \log f\left(\frac{t}{\beta_n}\right) - it\gamma_n = it\gamma - \sigma^2\frac{t^2}{2}$$

$$+ \int_{\Delta_n}\left(e^{itu} - 1 - \frac{itu}{1+u^2}\right)\frac{1+u^2}{u^2}\,dG(u) + o(1).$$

Therefore

$$q_n \log f\left(\frac{t}{\beta_n}\right) - it\gamma_n \Rightarrow$$

$$\Rightarrow it\gamma + \int_{-\infty}^{\infty}\left(e^{itu} - 1 - \frac{itu}{1+u^2}\right)\frac{1+u^2}{u^2}\,dG(u) = \log v(t).$$

This relation shows that as $n \to \infty$ the distribution functions of the sums

$$\zeta_{q_n} = \frac{\xi_1 + \xi_2 + \cdots + \xi_{q_n}}{\beta_n} - \gamma_n$$

of independent summands $\xi_k$, distributed according to the law $F(x)$, converge to $V(x)$. We have therefore found for every infinitely divisible law $V(x)$ a law $F(x)$ belonging to its domain of partial attraction, proving the theorem.

Soon after A. Ya. Khintchine proved this theorem examples of laws which do not belong to the domain of partial attraction of any proper law * were constructed simultaneously by three authors (P. Lévy [76], p. 212, A. Ya. Khintchine [59], B. V. Gnedenko [38]). The idea in the

* As is easily proved, improper laws attract all laws.

construction of all these examples was the same — the consideration of a random variable with sufficiently large probabilities for large values.

Consider the distribution function $F(x)$ for which

$$\log f(t) = \int_{-\infty}^{-e} (\cos tx - 1)\, d\left(\frac{1}{4\log|x|}\right) + \int_{e}^{\infty} (\cos tx - 1)\, d\left(-\frac{1}{4\log x}\right).$$

$F(x)$ as an infinitely divisible law has a non-empty domain of partial attraction, but, as we shall see now, is itself not attracted to any law except the improper ones.

We examine the behavior of the function $\log f(t)$ near the point $t = 0$. To this end we note the equation

$$\log f(t) = \int_{e}^{\infty} \sin^2 \frac{tx}{2}\, d\left(\frac{1}{\log x}\right)$$

$$= \int_{e}^{\frac{1}{\sqrt{|t|}}} \sin^2 \frac{tx}{2}\, d\left(\frac{1}{\log x}\right) + \int_{\frac{1}{\sqrt{|t|}}}^{\infty} \sin^2 \frac{tx}{2}\, d\left(\frac{1}{\log x}\right).$$

Since all the integrals considered are negative,

$$\int_{e}^{\frac{1}{\sqrt{|t|}}} \sin^2 \frac{tx}{2}\, d\left(\frac{1}{\log x}\right) > \int_{e}^{\frac{1}{\sqrt{|t|}}} \frac{t^2x^2}{4}\, d\left(\frac{1}{\log x}\right)$$

$$> \left(\frac{1}{\sqrt{|t|}}\right)^2 \frac{t^2}{4} \int_{e}^{\infty} d\left(\frac{1}{\log x}\right) = -\frac{|t|}{4}$$

and

$$\int_{\frac{1}{\sqrt{|t|}}}^{\infty} \sin^2 \frac{tx}{2}\, d\left(\frac{1}{\log x}\right) > \int_{\frac{1}{\sqrt{|t|}}}^{\infty} d\left(\frac{1}{\log x}\right) = \frac{2}{\log|t|}.$$

On the other hand, for sufficiently small $t$

$$\int_{\frac{1}{\sqrt{|t|}}}^{\infty} \sin^2 \frac{tx}{2}\, d\left(\frac{1}{\log x}\right) < \int_{\frac{\pi}{2|t|}}^{\infty} \sin^2 \frac{tx}{2}\, d\left(\frac{1}{\log x}\right) < \sum_{n=0}^{\infty} \int_{\frac{4n+1}{2|t|}\pi}^{\frac{4n+3}{2|t|}\pi} \sin^2 \frac{xt}{2}\, d\left(\frac{1}{\log x}\right)$$

$$< \sum_{n=0}^{\infty} \int_{\frac{4n+1}{2|t|}\pi}^{\frac{4n+3}{2|t|}\pi} \left(\frac{\sqrt{2}}{2}\right)^2 d\left(\frac{1}{\log x}\right).$$

Now

$$\int_{\frac{4n+1}{2|t|}\pi}^{\frac{4n+3}{2|t|}\pi} d\left(\frac{1}{\log x}\right) < \int_{\frac{4n+3}{2|t|}\pi}^{\frac{4n+5}{2|t|}\pi} d\left(\frac{1}{\log x}\right),$$

so that

$$\sum_{n=0}^{\infty} \int_{\frac{4n+1}{2|t|}\pi}^{\frac{4n+3}{2|t|}\pi} d\left(\frac{1}{\log x}\right) < \frac{1}{2} \int_{\frac{\pi}{2|t|}}^{\infty} d\left(\frac{1}{\log x}\right) = -\frac{1}{2\log\frac{\pi}{2|t|}} < \frac{1}{3\log|t|}$$

and consequently

$$\int_{\frac{1}{\sqrt{|t|}}}^{\infty} \sin^2\frac{tx}{2}\, d\frac{1}{\log x} < \frac{1}{6\log|t|}.$$

From the estimates obtained we conclude that for $t$ sufficiently near zero,

$$\log f(t) = \frac{A(t)}{\log|t|},$$

where $A(t)$ is a continuous function satisfying the inequalities

$$\frac{1}{10} < A(t) < 3.$$

We shall now prove that for every choice of the constants $B_n > 0$ [$A_n$ can be taken to be zero by the symmetry of the law $F(x)$] and natural numbers $n_1 < n_2 < \cdots < n_k < \cdots$ the function

$$n_k \log f\left(\frac{t}{B_{n_k}}\right)$$

cannot converge to a limit uniformly with respect to $t$ in a neighborhood of $t = 0$.

Consider all possible cases.

1. The sequences $\{B_n\}$ and $\{n_k\}$ are such that

$$\lim_{k\to\infty} \frac{\log B_{n_k}}{n_k} = a < +\infty.$$

In this case for every $t$ different from 0, the ratio $n_k/\log(|t|/B_{n_k})$ approaches $-(1/a)$; while for $t = 0$ it approaches 0.* This means that in the case considered the limit function for $n_k \log f(t/B_{n_k})$ cannot be the logarithm of a characteristic function.

2. The sequences $\{B_n\}$ and $\{n_k\}$ are such that

$$\lim_{k\to\infty} \frac{\log B_{n_k}}{n_k} = +\infty.$$

---

* *Translator's note.* What the authors mean is that $n_k \log f\left(\frac{t}{B_{n_k}}\right)$ is 0 for $t = 0$.

In this case, in any finite interval of $t$

$$\lim_{k\to\infty} n_k \log f\left(\frac{t}{B_{n_k}}\right) = 0$$

and consequently the limit function for $f^{n_k}(t/B_{n_k})$ will be the characteristic function of the unitary law.

3. If the ratio $\log B_{n_k}/n_k$ as $k \to \infty$ does not approach a limit, then we can choose a subsequence $n_k'$ from the sequence $n_k$ for which

$$\lim_{k\to\infty} \frac{\log B_{n_k}'}{n_k'} = a \leqslant +\infty,$$

and so reduce this last possible case to one of the previous cases.

We have thereby proved our assertion.

Doeblin [24], B. V. Gnedenko [39], A. V. Groshev [47], and P. Lévy [76] occupied themselves with the further investigation of domains of partial attractions. We shall cite some of the results without pausing for their proofs.

1. Each distribution law $F(x)$ belongs to the domain of partial attraction of one or a nondenumerable set of types or else does not belong to any domain of partial attraction at all (Doeblin [24], B. V. Gnedenko [39]).

2. If a distribution law belongs to the domain of partial attraction of only one type, then this type must be stable (Gnedenko [39]).

3. The domain of partial attraction of a stable type is wider than its domain of (complete) attraction (Gnedenko [39], Doeblin [24]).

4. If the law $F(x)$ belongs to the domain of partial attraction of the law $V(x)$, and the law $V(x)$ belongs to the domain of partial attraction of the law $\Psi(x)$, then $F(x)$ belongs to the domain of partial attraction of the law $\Psi(x)$ (B. V. Gnedenko [39]).

5. From the result 4 it follows in particular that every law which belongs to the domain of partial attraction of a type with finite variance, belongs also to the domain of partial attraction of the normal type; that the only stable type with finite variance is the normal type; that the result 3 is true.*

6. There exist laws $F(x)$ belonging to the domain of partial attraction of every infinitely divisible type (the universal laws according to Doeblin's terminology [24]).

---

* *Translator's note.* To see that result 3 follows from 4, we need the following fact. If $L$ is a stable law then there exists an infinitely divisible law $L_1$ of a different type from $L$ which belongs to the domain of attraction of $L$. If $L$ is normal this is a consequence of the remark at the end of § 30 and Theorem 4 of § 35. If $L$ is stable of exponent $\alpha < 2$, let $N$ be the normal law with mean 0 and variance 1. By Theorem 5 of § 35, $L_1 = L * N$ is in the domain of normal attraction of $L$. Now let $L_2$ be partially attracted to $L_1$. It follows from 4 that $L_2$ is partially attracted to $L$ but $L_2$ cannot be in the domain of attraction of $L$ because it is partially attracted to $L_1$. Hence 3 is true.

7. In order that the law $F(x)$ belong to no domain of partial attraction, it is sufficient that (Doeblin [24])

$$\lim_{l \to \infty} \varliminf_{X \to \infty} \frac{\int\limits_{|x| > lX} dF(x)}{\int\limits_{|x| > X} dF(x)} \neq 0.$$

8. In order that the law $F(x)$ belong to the domain of partial attraction of the normal type, it is necessary and sufficient that (P. Lévy [76], p. 113)

$$\varliminf_{X \to \infty} \frac{X^2 \int\limits_{|x| > X} dF(x)}{\int\limits_{|x| < X} x^2\, dF(x)} = 0. \tag{5}$$

9. In order that the law $F(x)$ belong to the domain of partial attraction of the Poisson law, it is necessary and sufficient (A. V. Groshev [47]) that for every $\epsilon > 0$

$$\varliminf_{l \to \infty} \frac{\int\limits_{|x-1| > \epsilon} \frac{x^2}{1+x^2}\, dF(lx)}{\int\limits_{|x-1| < \epsilon} dF(lx)} = 0. \tag{6}$$

According to result 5 just cited, every law belonging to the domain of partial attraction of the Poisson law belongs also to the domain of partial attraction of the normal type; consequently if equation (6) is satisfied, so also is (5).

## CHAPTER 8

## IMPROVEMENT OF THEOREMS ABOUT THE CONVERGENCE TO THE NORMAL LAW

### § 38. Statement of the Problem

The present chapter is devoted exclusively to the convergence of normalized sums to the normal law. In this connection we confine ourselves to the case of identically distributed summands. Moreover, we require the existence of moments for the summands: in all cases the second moment, and in a number of cases also moments of higher order.

Some of the theorems presented here can be proved also under assumptions different from those we make, in particular, for nonidentically distributed summands.

The basic problem, in which we are interested here, consists in the study of the asymptotic behavior of the difference between the distribution function $F_n(x)$ of the normalized sum of the first $n$ terms of the sequence of independent random variables

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

and the normal distribution function $\Phi(x)$.

Throughout this chapter we shall suppose that

$$\mathsf{M}\,\xi_n = 0.$$

This restriction, of course, does not diminish the generality of our considerations. The principle of the solution of the problem stated was indicated by Chebyshev in his fundamental paper of the year 1887 * [17], in which he gives the following expansion for the difference $F_n(x) - \Phi(x)$:

$$F_n(x) - \Phi(x) \sim \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \left( \frac{Q_1(x)}{n^{\frac{1}{2}}} + \frac{Q_2(x)}{n} + \cdots + \frac{Q_j(x)}{n^{\frac{j}{2}}} + \cdots \right), \quad (1)$$

where the $Q_j(x)$ are polynomials, the coefficients of which depend only on the first $j + 2$ moments of the random variable $\xi_n$.

At the basis of the expansion (1) lies the more general idea, also due to Chebyshev, of expanding an arbitrary function $p(x)$ in a series of Chebyshev-Hermite polynomials. These polynomials were introduced by Chebyshev in 1859 [18]. We call them Chebyshev-Hermite polynomials. The name of Hermite, who discovered them much later, is adjoined to

* *Translator's note.* The year given is the year when the paper was first published in Russian, hence it is not the same as that given in the Literature.

that of Chebyshev solely for the sake of distinguishing these from the great number of other important types of polynomials which justly bear or ought to bear the name Chebyshev.

The Chebyshev-Hermite polynomials can be defined by the formula

$$H_k(x) = (-1)^k e^{\frac{x^2}{2}} \frac{d^k}{dx^k} e^{-\frac{x^2}{2}}, \tag{2}$$

which gives

$$\left.\begin{aligned} &H_0(x) = 1, \\ &H_1(x) = x, \\ &H_2(x) = x^2 - 1, \\ &H_3(x) = x^3 - 3x, \\ &\dots\dots\dots\dots \end{aligned}\right\} \tag{3}$$

The expansion of an arbitrary function proposed by Chebyshev in 1859 has the form

$$p(x) \sim \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{c_k}{k!} \frac{d^k}{dx^k} e^{-\frac{x^2}{2}} \sim \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} (-1)^k \frac{c_k}{k!} e^{-\frac{x^2}{2}} H_k(x), \tag{4}$$

where

$$c_k = (-1)^k \int H_k(x) p(x)\, dx. \tag{5}$$

Since

$$H_k(x) = \sum_{j=0}^{k} h_{jk} x^j \tag{6}$$

is a polynomial of the $k$th degree, the coefficients $c_k$ can be expressed in the form

$$c_k = (-1)^k \sum_{j=0}^{k} h_{jk} \alpha_j \tag{7}$$

by means of the moments

$$\alpha_j = \int p(x) x^j\, dx; \qquad j = 0, 1, 2, \dots, k. \tag{8}$$

In the case of the probability density $p_\xi(x)$ of a normalized random variable $\xi$ with

$$\mathsf{M}\xi = 0, \qquad \mathsf{D}^2\xi = 1 \tag{9}$$

the first few moments have the values

$$\alpha_0 = 1, \quad \alpha_1 = 0, \quad \alpha_2 = 1,$$

and hence the expansion (4) takes the simplified form

$$p_\xi(x) \sim \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(1 - \frac{c_3}{6} H_3(x) + \frac{c_4}{24} H_4(x) - \dots\right). \tag{10}$$

Put

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}}\, dx. \tag{11}$$

Then for the distribution function $F_\xi(x)$ it is natural to consider the integrated expansion (10):

$$F_\xi(x) \sim \Phi(x) + \frac{c_3}{6}\Phi^{(3)}(x) + \frac{c_4}{24}\Phi^{(4)}(x) + \ldots, \tag{12}$$

where

$$\Phi^{(k)}(x) = \frac{(-1)^{k-1}}{\sqrt{2\pi}} H_{k-1}(x)\, e^{-\frac{x^2}{2}}.$$

By successive integrations by parts it is not difficult to verify the equation

$$\int e^{itx}\, d\Phi^{(k)}(x) = \frac{(-it)^k}{\sqrt{2\pi}} \int e^{itx - \frac{x^2}{2}} dx = (-it)^k e^{-\frac{t^2}{2}}. \tag{13}$$

Thus to the expansion (12) corresponds formally the expansion of the characteristic function

$$f_\xi(t) = e^{-\frac{t^2}{2}} \left[ 1 + \sum_{k=3}^{\infty} \frac{c_k}{k!}(-it)^k \right]. \tag{14}$$

From this we can deduce expressions for the coefficients $c_r$ by means of semi-invariants. For this purpose we make use of the formal relation (see § 15)

$$\log f_\xi(t) = \sum_{r=1}^{\infty} \frac{\varkappa_r}{r!}(it)^r.$$

In our case of a normalized random variable,

$$\varkappa_1 = 0, \qquad \varkappa_2 = 1,$$

which gives

$$\log f_\xi(t) = -\frac{t^2}{2} + \sum_{r=3}^{\infty} \frac{\varkappa_r}{r!}(it)^r. \tag{15}$$

Putting $w = -it$, we obtain from (14) and (15)

$$1 + \sum_{r=3}^{\infty} \frac{c_r}{r!} w^r = e^{\sum_{r=3}^{\infty} \frac{\varkappa_r}{r!}(-w)^r}, \tag{16}$$

i.e.,

$$\left.\begin{aligned} c_3 &= -\varkappa_3, \\ c_4 &= \varkappa_4, \\ c_5 &= -\varkappa_5, \\ c_6 &= \varkappa_6 + 10\varkappa_3^2, \\ &\cdots\cdots\cdots \end{aligned}\right\} \tag{17}$$

It is easy to see that the coefficient $c_r$ depends only on the first $r$ semi-invariants.

After these general remarks we now turn to the case of the distribution functions $F_n(x)$ of the normalized sums

$$\zeta_n = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{\sigma \sqrt{n}}$$

of identically distributed summands $\xi_k$ with

$$\mathbf{M}\xi_k = 0, \quad \mathbf{D}^2\xi_k = \sigma^2.$$

Putting

$$\xi'_k = \frac{\xi_k}{\sigma},$$

we can write $\zeta_n$ in the form

$$\zeta_n = \frac{\xi'_1 + \xi'_2 + \ldots + \xi'_n}{\sqrt{n}}.$$

Denoting the semi-invariants of $\xi_k$ by $\chi_k$, we easily find (see § 15) that the semi-invariants of $\xi'_k$ are

$$\lambda_1 = 0, \quad \lambda_2 = 1, \quad \lambda_r = \frac{\varkappa_r}{\sigma^r} \quad \text{for} \quad r > 2, \tag{18}$$

and the semi-invariants of $\zeta_n$ are

$$\varkappa_1^{(n)} = 0, \quad \varkappa_2^{(n)} = 1,$$

$$\varkappa_r^{(n)} = \frac{\lambda_r}{n^{\frac{r-2}{2}}} = \frac{\varkappa_r}{n^{\frac{r-2}{2}}\sigma^r} \quad \text{for } r > 2. \tag{19}$$

The expansion (12) in our case takes the form

$$F_n(x) - \Phi(x) \sim \sum_{r=3}^{\infty} \frac{c_r^{(n)}}{r!} \Phi^{(r)}(x), \tag{20}$$

where the coefficients $c_r^{(n)}$ are calculated from the formal equation

$$1 + \sum_{r=3}^{\infty} \frac{c_r^{(n)}}{r!} w^r = \exp\left\{ \sum_{r=3}^{\infty} \frac{\lambda_r}{r!\, n^{(r-2)/2}} (-w)^r \right\}, \tag{21}$$

which gives

$$\left.\begin{aligned} c_3^{(n)} &= -\frac{\lambda_3}{n^{\frac{1}{2}}}, \\ c_4^{(n)} &= \frac{\lambda_4}{n}, \\ c_5^{(n)} &= -\frac{\lambda_5}{n^{\frac{3}{2}}}, \\ c_6^{(n)} &= \frac{\lambda_6}{n^2} + \frac{10\lambda_3^2}{n}, \\ & \dots\dots\dots\dots \end{aligned}\right\} \tag{22}$$

After substituting the expressions for $c_r^n$ into (20), it is natural to collect terms of the same order in $n$. This then leads to Chebyshev's expansion

$$F_n(x) - \Phi(x) \sim \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \left( \frac{Q_1(x)}{n^{\frac{1}{2}}} + \frac{Q_2(x)}{n} + \frac{Q_3(x)}{n^{\frac{3}{2}}} + \dots \right). \tag{23}$$

It is easily seen that

$$\left.\begin{aligned} Q_1(x) &= \frac{\lambda_3}{6}(1 - x^2), \\ Q_2(x) &= -\frac{10\lambda_3^2}{6!} x^5 + \frac{1}{8}\left( \frac{\lambda_4}{3} - \frac{10\lambda_3^2}{9} \right) x^3 + \left( \frac{5\lambda_3^2}{24} - \frac{\lambda_4}{8} \right) x, \\ & \dots\dots\dots\dots\dots\dots\dots\dots \end{aligned}\right\} \tag{24}$$

A general method of calculating the polynomials $Q_r$ is contained in the following. Expand the right side of (21) in powers of $1/\sqrt{n}$:

$$e^{\sum\limits_{k=1}^{\infty} \frac{\lambda_{k+2}}{(k+2)!} (-w)^{k+2} \left( \frac{1}{\sqrt{n}} \right)^k} = 1 + \sum_1^{\infty} P_k(-w) \left( \frac{1}{\sqrt{n}} \right)^k. \tag{25}$$

It is easily seen that $P_k(-w)$ is a polynomial in $w$ of degree $3k$ with coefficients depending on $\lambda_3, \lambda_4, \dots, \lambda_{k+2}$:

$$\left.\begin{aligned} P_1(-w) &= \frac{\lambda_3}{6}(-w)^3, \\ P_2(-w) &= \frac{\lambda_4}{24} w^4 + \frac{\lambda_3^2}{72} w^6, \\ & \dots\dots\dots\dots\dots \end{aligned}\right\} \tag{26}$$

Comparing the expansion

$$1 + \sum_{r=3}^{\infty} \frac{c_r^{(n)}}{r!} w^r = 1 + \sum_1^{\infty} P_k(-w) \left( \frac{1}{\sqrt{n}} \right)^k \tag{27}$$

with the expansion (20), we obtain

$$F_n(x) - \Phi(x) \sim \sum_{1}^{\infty} P_k(-\Phi)\left(\frac{1}{\sqrt{n}}\right)^k, \tag{28}$$

where $P_k(-\Phi)$ is calculated by replacing $w^r$ by $\Phi^{(r)}$ in $P_k(-w)$.

From (1) and (28) we deduce a general formula for finding the polynomials $Q_k$:

$$\frac{1}{\sqrt{2\pi}} Q_k(x) e^{-\frac{x^2}{2}} = P_k(-\Phi). \tag{29}$$

Expansion of the type (12) was studied after Chebyshev by Bruns [13] and Charlier [15] from the standpoint of probability. Chebyshev's expansion in the form (28) was studied in detail by Edgeworth [25].

The central problem of this chapter consists in the study of the asymptotic behavior as $n \to \infty$ of the remainder term

$$R_k^{(n)}(x) = F_n(x) - \Phi(x) - \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\left[\frac{Q_1(x)}{\sqrt{n}} + \ldots + \frac{Q_k(x)}{(\sqrt{n})^k}\right]$$

of Chebyshev's expansion. The most definitive result in this direction, due to Cramér, is presented in § 45. The case $k = 1$ is studied in detail in §§ 42–43, where results are obtained under broader assumptions than in § 45. In § 47 is given a theorem analogous to the theorem in § 45 but for probability densities. § 46 has a more elementary character: there we discuss the question of conditions for the convergence of the probability densities $p_n(x)$ of normalized sums $\zeta_n$ to the normal density

$$\varphi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

without regard to further improvements and estimates of the remainder terms. §§ 39 and 41 are auxiliary in character, and § 40 contains a result of which the complete elucidation is given in § 42.

## § 39. Two Auxiliary Theorems

THEOREM 1. *Let $A$, $T$, and $\epsilon > 0$ be constants, $F(x)$ a nondecreasing function, and $G(x)$ a function of bounded variation. If*

1. $F(-\infty) = G(-\infty), \quad F(+\infty) = G(+\infty)$;

2. $\int |F(x) - G(x)|\, dx < \infty$;

3. $G'(x)$ exists for all $x$ and $|G'(x)| \leq A$;

4. $\int_{-T}^{T} \left|\frac{f(t) - g(t)}{t}\right| dt = \varepsilon,$

*then to every number $k > 1$ there corresponds a finite positive number $c(k)$ depending only on $k$ such that*

$$|F(x)-G(x)| \leqslant k\frac{\varepsilon}{2\pi}+c(k)\frac{A}{T}. \tag{1}$$

*Proof.* First of all we note that from the equation

$$f(t)-g(t)=\int e^{itx}d\,[F(x)-G(x)]$$

we can deduce by integration by parts that *

$$\frac{f(t)-g(t)}{-it}=\int e^{itx}\,[F(x)-G(x)]\,dx. \tag{2}$$

We note further that for the proof of the theorem it is sufficient to consider the case $A = T = 1$. Otherwise, indeed, we may consider the functions

$$F_1(x)=\frac{T}{A}F\left(\frac{x}{T}\right)$$

and

$$G_1(x)=\frac{T}{A}G\left(\frac{x}{T}\right).$$

Evidently,

$$|G_1'(x)|\leqslant 1,\quad \int_{-1}^{1}\left|\frac{f_1(t)-g_1(t)}{t}\right|dt=\frac{T\varepsilon}{A}=\varepsilon_1.$$

If we suppose that the theorem is proved for the case $A = T = 1$, then

$$|F_1(x)-G_1(x)|\leqslant k\frac{\varepsilon_1}{2\pi}+c(k).$$

Using the definition of $F_1(x)$, $G_1(x)$, and $\varepsilon_1$, we obtain (1).

Consider the functions

$$H(x)=\frac{3}{8\pi}\left(\frac{\sin\frac{x}{4}}{\frac{x}{4}}\right)^4$$

and

$$h(t)=\begin{cases} 0 & \text{for } |t|\geqslant 1,\\ 2(1-|t|)^3 & \text{for } \frac{1}{2}\leqslant|t|\leqslant 1,\\ 1-6t^2+6|t|^3 & \text{for } 0\leqslant|t|\leqslant\frac{1}{2}. \end{cases}$$

* *Translator's note.* For $t = 0$ the left side of (2) is defined as the limit as $t \to 0$, which is finite by the assumption 2.

It is easily verified that they satisfy the following relations:

1. $h(t) = \int e^{itx} H(x)\,dx,$

2. $\int |x| H(x)\,dx = b < \infty,$

3. $\int H(x)\,dx = 1.$

Construct the function

$$v(x) = \int H(x-y)\,[F(y) - G(y)]\,dy.$$

According to (2) together with Theorem 3 of § 11, the equation

$$\frac{f(t)-g(t)}{-it}\,h(t) = \int e^{itx} v(x)\,dx$$

holds. Since the function on the left is absolutely integrable,

$$v(x) = \frac{1}{2\pi}\int e^{-itx}\,\frac{f(t)-g(t)}{-it}\,h(t)\,dt.$$

Using the definition of the function $v(x)$ and the fact that $h(t) = 0$ for $|t| \geq 1$, we find that

$$\int H(x-y)\,[F(y)-G(y)]\,dy = \frac{1}{2\pi}\int_{-1}^{1} e^{-itx}\,\frac{f(t)-g(t)}{-it}\,h(t)\,dt. \qquad (3)$$

We now put

$$\Delta = \max_{-\infty < x < \infty} |F(x) - G(x)|.$$

Without loss of generality we may suppose that

$$\Delta = |F(0) - G(0)|$$

and that $F(0) > G(0)$. Under these conditions, taking into account the relation $|G'(x)| \leq 1$, we find that for $0 \leq x \leq \Delta$

$$\Delta \geqslant F(x) - G(x) \geqslant \Delta - |x|. \qquad (4)$$

From (3) we conclude that

$$\left|\int H(x-y)\,[F(y)-G(y)]\,dy\right| \leqslant \frac{1}{2\pi}\int_{-1}^{1}\left|\frac{f(t)-g(t)}{t}\right|\cdot\left|h(t)\right|dt$$

$$\leqslant \frac{1}{2\pi}\int_{-1}^{1}\left|\frac{f(t)-g(t)}{t}\right|dt = \frac{\varepsilon}{2\pi}.$$

By (4) it is obvious that

$$\frac{\varepsilon}{2\pi} \geqslant \left|\int H(x-y)\,[G(y)-F(y)]\,dy\right|$$

$$\geqslant \int_0^{\Delta} (\Delta - |y|)\,H(x-y)\,dy - \int_{-\infty}^{0} H(x-y)\,|F(y)-G(y)|\,dy -$$

(cont'd)

$$-\int_{\Delta}^{\infty} H(x-y)\,|F(y)-G(y)|\,dy$$

$$\geqslant \int_{0}^{\Delta} (\Delta-y)\,H(x-y)\,dy-\Delta\left[\int_{-\infty}^{0} H(x-y)\,dy\right.$$

$$\left.+\int_{\Delta}^{\infty} H(x-y)\,dy\right]=\int_{0}^{\Delta} (2\Delta-y)\,H(x-y)\,dy-\Delta.$$

But

$$\int_{0}^{\Delta} (2\Delta-y)\,H(x-y)\,dy=\int_{-x}^{\Delta-x} (2\Delta-x-z)\,H(z)\,dz$$

$$\geqslant (2\Delta-x)\int_{-x}^{\Delta-x} H(z)\,dz-\int |z|\,H(z)\,dz,$$

so that

$$-\Delta+(2\Delta-x)\int_{-x}^{\Delta-x} H(z)\,dz \leqslant \frac{\varepsilon}{2\pi}+b.$$

Since $0 < x < \Delta$, we can find a $\theta$ $(0 < \theta < 1)$ such that $x = \Delta\theta$. Thus the preceding inequality takes the form

$$\Delta\left\{(2-\theta)\int_{-\theta\Delta}^{\Delta(1-\theta)} H(z)\,dz-1\right\} \leqslant \frac{\varepsilon}{2\pi}+b.$$

Whatever $k > 1$ may be, it is always possible to choose $\theta(k) > 0$ and $\alpha(k) > 0$, so that

$$(2-\theta(k))\int_{-\theta(k)\,\alpha(k)}^{(1-\theta(k))\,\alpha(k)} H(z)\,dz-1=\frac{1}{k}. \tag{5}$$

Consider the two possible cases:

$$1.\ \Delta \leqslant \alpha(k), \qquad 2.\ \Delta > \alpha(k).$$

In the second case we choose $x$ so that $\theta = \theta(k)$. It is then evident that

$$1+\frac{1}{k}=(2-\theta(k))\int_{-\alpha(k)\,\theta(k)}^{(1-\theta(k))\,\alpha(k)} H(z)\,dz \leqslant (2-\theta)\int_{-\Delta\theta}^{\Delta(1-\theta)} H(z)\,dz.$$

Thus in this case,

$$\Delta \leqslant \frac{k\varepsilon}{2\pi}+kb < \frac{k\varepsilon}{2\pi}+(kb+\alpha(k)).$$

In the first case,

$$\Delta \leqslant \alpha(k) < \frac{k\varepsilon}{2\pi} + (kb + \alpha(k)).$$

Putting $c(k) = kb + \alpha(k)$, we complete the proof of the theorem.

We need another theorem, similar to the one just proved, but in which it is not assumed that the function $G(x)$ is continuous for all values $x$.

THEOREM 2. *Let $A$, $T$, $\epsilon$ be arbitrary positive constants, $F(x)$ a nondecreasing purely discontinuous function, and $G(x)$ a function of bounded variation. If*

1) $F(-\infty) = G(-\infty) = 0, \quad F(+\infty) = G(+\infty),$

2) $\int |F(x) - G(x)|\, dx < \infty,$

3) *the functions $F(x)$ and $G(x)$ have discontinuities only at the points $x = x_v$ ($x_v < x_{v+1}$; $v = 0, \pm 1, \pm 2, \ldots$), and there exists an $l$ such that* $\min (x_{v+1} - x_v) \geq l$,

4) *everywhere except at $x = x_v$ ($v = 0, \pm 1, \pm 2, \ldots$),*

$$|G'(x)| \leqslant A,$$

5) $\displaystyle\int_{-T}^{T} \left|\frac{f(t) - g(t)}{t}\right| dt = \varepsilon,$

*then to every number $k > 1$ there correspond two finite numbers $c_1(k)$ and $c_2(k)$ depending only on $k$ and such that*

$$|F(x) - G(x)| \leqslant k\frac{\varepsilon}{2\pi} + c_1(k)\frac{A}{T},$$

*whenever $T \cdot l \geq c_2(k)$.*

*Proof*. As in the theorem just proved, we may put $A = T = 1$ and

$$\Delta = \sup_{-\infty < x < +\infty} |F(x) - G(x)| = |F(0) - G(0)|.$$

The behavior of the functions $F(x)$ and $G(x)$ in the neighborhood of $x = 0$ can be reduced to several cases, each of which must be examined individually. We confine ourselves to the case that $x_v \neq 0$ for every $v$ and the distance from the origin to the nearest $x_v > 0$ is not less than $l/2$. If we put $\delta = \min (\Delta, l/2)$, then, as in the preceding theorem, we find that (for $x = \delta\theta$)

$$\Delta \left\{(2 - \theta) \int_{-\theta\delta}^{(1-\theta)\delta} H(y)\, dy - 1\right\} \leqslant \frac{\varepsilon}{2\pi} + b.$$

For an arbitrary $k > 1$ we can again choose a sufficiently small $\theta(k)$ and a sufficiently large $\alpha(k)$ so that (5) holds. If $\Delta \geq \alpha(k)$ and $l/2 \geq \alpha(k)$, then $\delta \geq \alpha(k)$ and therefore

$$\Delta \leqslant k\frac{\varepsilon}{2\pi} + kb < k\frac{\varepsilon}{2\pi} + (kb + \alpha(k)).$$

In case $\Delta < \alpha(k)$ this inequality is obviously satisfied. Putting $c_1 = kb + \alpha(k)$ and $c_2(k) = 2\alpha(k)$, we obtain the proof of the theorem.

## § 40. Estimation of the Remainder Term in Lyapunov's Theorem

Before turning to the essence of the problem, we shall introduce some notations which will be adhered to in what follows. We put

$$\rho_s = \frac{\beta_s}{\beta_2^{\frac{s}{2}}} = \frac{\beta_s}{\sigma^s} \text{ and } \lambda_s = \frac{\varkappa_s}{\sigma^s}. \tag{1}$$

Obviously, $\rho_s$ and $\lambda_s$ are respectively the $s$th moment and the $s$th semi-invariant of the random variable $\xi/\sigma$. Therefore, according to (5) and (2) of § 15, the following inequalities are satisfied:

$$0 \leqslant \rho_1 \leqslant \rho_2^{\frac{1}{2}} \leqslant \rho_3^{\frac{1}{3}} \leqslant \ldots \leqslant \rho_s^{\frac{1}{s}} \leqslant \ldots, \tag{2}$$

$$|\lambda_s| < s^s \rho^s. \tag{3}$$

Lyapunov in the proof of his theorem obtained not only the proposition about the convergence of the distribution functions of the normalized sums to the normal law, but also an estimate of the speed of convergence. In case the existence of the third moments is assumed, the inequality

$$|F_n(x) - \Phi(x)| < c\rho_3 \frac{\lg n}{\sqrt{n}},$$

holds, where $c$ is a constant (H. Cramér [19] proved that $c$ can be taken to be 3).

The object of this section is to prove the following more definitive result, first obtained by H. Cramér [21] under some additional assumptions and in the form cited here by Esseen [26] and A. C. Berry [10]. (For the extension of this proposition to the case of nonidentically distributed summands see Cramér [21].)

Theorem 1. *If the random variables $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ have finite third moments, then*

$$|F_n(x) - \Phi(x)| \leqslant c\frac{\rho_3}{\sqrt{n}},$$

*where $c$ is a constant* (in the paper [10] Berry proved that $c$ can be taken to be 1.88 *).

We remark that the order of the estimate, which is given by this theorem, cannot be improved in the general case, even if the existence of

* *Translator's note.* Berry's computation is invalidated by an error; see P. L. Hsu, The Approximate distributions of the mean and variance of a sample of independent variables, *Annals of Mathematical Statistics*, **14**, 1–29 (1945).

moments of all orders is assumed for the summands. It is easy to convince ourselves of this by considering, for example, summands which are identically distributed and take only two values: $-1$ and $+1$, each with probability $\frac{1}{2}$. At the point $x = 0$ the function $F_n(x)$, as it follows from the local theorem of de Moivre-Laplace, has a jump asymptotically equal to $1/\sqrt{2\pi n}$. This obviously proves the remark just made.

The proof of the theorem formulated above is based on Theorem 1 of § 39 and the following proposition.

THEOREM 2. *If the random variables $\xi_1, \xi_2, \ldots, \xi_n$ are identically distributed and have finite third moments, then for*

$$|t| \leqslant \frac{\sigma^3 \sqrt{n}}{5\beta_3} = T_n$$

*the following inequality holds:*

$$\left| f_n(t) - e^{-\frac{t^2}{2}} \right| \leqslant \frac{7}{6} \frac{|t|^3 \beta_3}{\sigma^3 \sqrt{n}} e^{-\frac{t^2}{4}}.$$

*Proof*. Indeed, from

$$f\left(\frac{t}{B_n}\right)^* = 1 - \frac{t^2}{2n} + i \frac{t^3}{6B_n^3} \int e^{i\theta t x} x^3 \, dF(x),$$

we deduce that

$$\left| f\left(\frac{t}{B_n}\right) \right| \geqslant 1 - \frac{t^2}{2n} - \frac{|t|^3}{6B_n^3} \beta_3$$

$$\geqslant 1 - \frac{T_n^2}{2n} - \frac{T_n^3}{6B_n^3} \beta_3 = 1 - \frac{\sigma^6}{50\beta_3^2}\left(1 + \frac{1}{15}\right).$$

Since according to (5) of § 15 $\sigma^6 = \beta_2^3 \leq \beta_3^2$,

$$\left| f\left(\frac{t}{B_n}\right) \right| > \frac{24}{25}$$

and consequently $f\left(\frac{t}{B_n}\right)$ is different from zero for $|t| \leq T_n$. Therefore in this interval we may write

$$f_n(t) = e^{n \log f\left(\frac{t}{B_n}\right)}.$$

But

$$n \log f\left(\frac{t}{B_n}\right) = -\frac{t^2}{2} + \frac{nt^3 \alpha}{6B_n^3},$$

where

$$\alpha = \left[\frac{d^3}{dz^3} \log f(z)\right]_{z = \theta \frac{t}{B_n}}.$$

* *Translator's note.* $B_n = \sqrt{n\beta_2} = \sqrt{n}\sigma$.

Consequently,

$$\left|f_n(t)-e^{-\frac{t^2}{2}}\right|=e^{-\frac{t^2}{2}}\left|e^{\frac{t^3\alpha}{6\sigma^3\sqrt{n}}}-1\right|.$$

Now

$$|e^{\beta}-1|=\left|\sum_{k=1}^{\infty}\frac{\beta^k}{k!}\right|\leqslant\sum_{k=1}^{\infty}\frac{|\beta|^k}{(k-1)!}=|\beta|e^{|\beta|}$$

and

$$\left|\frac{d^3}{dz^3}\log f(z)\right|\leqslant 7\beta_3.^*$$

Hence

$$|f_n(t)-e^{-\frac{t^2}{2}}|\leqslant e^{-\frac{t^2}{2}}\cdot\frac{7|t|^3\beta_3}{6\sigma^3\sqrt{n}}\cdot e^{\frac{7|t|^3\beta_3}{6\sigma^3\sqrt{n}}}.$$

Finally,

$$\frac{t^2}{4}-\frac{7|t|^3\beta_3}{6\sigma^3\sqrt{n}}=\frac{t^2}{4}\left(1-\frac{14}{3}\frac{|t|}{\sqrt{n}}\cdot\frac{\beta_3}{\sigma^3}\right)\geqslant\frac{t^2}{4}\left(1-\frac{5T_n}{\sqrt{n}}\cdot\frac{\beta_3}{\sigma^3}\right)=0,$$

hence

$$|f_n(t)-e^{-\frac{t^2}{2}}|\leqslant\frac{7}{6}\frac{|t|^3\beta_3}{\sigma^3\sqrt{n}}e^{-\frac{t^2}{4}}.$$

Now the proof of Theorem 1 can be obtained in a few words. In fact, we put in Theorem 1 of § 39

$$F(x)=F_n(x),\quad G(x)=\Phi(x),$$

$$A=\max_{-\infty<x<+\infty}|\Phi'(x)|=\frac{1}{\sqrt{2\pi}},\quad T=T_n=\frac{\sqrt{n}}{5\rho_3}.$$

---

* This estimate can be obtained as follows:

$$\left|\frac{d^3}{dz^3}\log f(z)\right|=\left|\frac{(f'''\cdot f-f'\cdot f'')\cdot f-2f'(f''\cdot f-f'^2)}{f^3}\right|$$

$$\leqslant\frac{\beta_3+3\beta_1\beta_2+2\beta_1^3}{|f|^3}\leqslant\frac{\beta_3+3\beta_1\beta_2+2\beta_1^3}{\left(\frac{24}{25}\right)^3}.$$

Now

$$\beta_1^3\leqslant\beta_2^{\frac{3}{2}}\leqslant\beta_3,$$

hence

$$\beta_3+3\beta_1\beta_2+2\beta_1^3\leqslant 6\beta_3$$

and consequently

$$\left|\frac{d^3}{dz^3}\log f(z)\right|\leqslant 7\beta_3.$$

According to the proposition just proved,

$$\varepsilon = \int_{-T}^{T} \left| \frac{f_n(t) - e^{-\frac{t^2}{2}}}{t} \right| dt \leqslant \frac{7\rho_3}{6\sqrt{n}} \int t^2 e^{-\frac{t^2}{4}} dt = \frac{7}{3}\sqrt{\pi}\, \frac{\rho_3}{\sqrt{n}}.$$

Therefore by Theorem 1 of § 39

$$|F_n(x) - \Phi(x)| \leqslant \frac{k}{2\pi} \frac{7\sqrt{\pi}}{3} \frac{\rho_3}{\sqrt{n}} + \frac{c(k)}{\sqrt{2\pi}} \cdot \frac{5\rho_3}{\sqrt{n}} = c \cdot \frac{\rho_3}{\sqrt{n}}.$$

## § 41. An Auxiliary Theorem

In § 38 we obtained for the characteristic function of the sum

$$\zeta_n = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{\sqrt{n}\sigma} \tag{1}$$

of independent identically distributed summands the formal expansion

$$f_n(t) \sim e^{-\frac{t^2}{2}} \left(1 + \sum_{k=1}^{\infty} P_k(it) \left(\frac{1}{\sqrt{n}}\right)^k\right).$$

In this section we deduce several properties of this expansion.

THEOREM 1. *If in the sum* (1) *the summands have finite moments up to the sth order inclusive* ($s \geq 3$), *then for* $|t| \leq T_{sn}{}^* = \dfrac{\sqrt{n}}{8s\rho_s^{3/s}}$ *the inequality*

$$\text{(a)} \quad \left|f_n(t) - e^{-\frac{t^2}{2}}\left(1 + \sum_{k=1}^{s-3} P_k(it)\left(\frac{1}{\sqrt{n}}\right)^k\right)\right|$$
$$\leqslant \frac{c_1(s)}{T_{sn}^{s-2}} \cdot \left(|t|^s + |t|^{3(s-2)}\right) e^{-\frac{t^2}{4}} \tag{2}$$

*holds, where* $c_1(s)$ *depends only on* $s$; *also, the inequality*

$$\text{(b)} \quad \left|f_n(t) - e^{-\frac{t^2}{2}}\left(1 + \sum_{k=1}^{s-2} P_k(it)\left(\frac{1}{\sqrt{n}}\right)^k\right)\right|$$
$$\leqslant c_2(s) \frac{\delta(n)}{T_{sn}^{s-2}} \left(|t|^s + |t|^{3(s-2)}\right) e^{-\frac{t^2}{4}} \tag{3}$$

*holds, where* $\delta(n)$ *depends only on* $n$ *and* $\lim\limits_{n\to\infty} \delta(n) = 0$.

*Proof of Theorem* 1(*a*).

By the hypothesis of the theorem, we can write

$$U = f\left(\frac{t}{B_n}\right) - 1 = \sum_{k=2}^{s-1} \frac{\alpha_k}{k!} \left(\frac{it}{B_n}\right)^k + \vartheta_1 \frac{\beta_s}{s!} \left(\frac{t}{B_n}\right)^s, \tag{4}$$

---

* *Translator's note.* The subscripts $s$ and $n$ are separate indices; $T_{sn}$ is not to be confused with the $T_n$ in § 40.

where $|\vartheta_1| \leq 1$. But for $|t| \leq T_{sn}$, according to (5) of § 15,

$$\frac{\beta_k^{\frac{1}{k}}|t|}{B_n} \leqslant \frac{\beta_s^{\frac{1}{s}} T_{sn}}{B_n} = \frac{1}{8s}\left(\frac{\beta_2^s}{\beta_s^2}\right)^{\frac{1}{s}} \leqslant \frac{1}{8s}.$$

Thus from the definition of $\alpha_k$ and $\beta_k$ and also (5) of § 15 it follows that

$$|U| \leqslant \sum_{k=2}^{\infty} \frac{1}{k!}\left(\frac{\beta_s^{\frac{1}{s}}|t|}{B_n}\right)^k \leqslant \sum_{k=2}^{\infty} \frac{1}{k!}\left(\frac{1}{8s}\right)^k \leqslant \frac{1}{100}.$$

Consequently,

$$\log f\left(\frac{t}{B_n}\right) = \log(1+U) = \sum_{1 \leqslant j < \frac{s}{2}} (-1)^{j+1}\frac{U^j}{j} + \frac{2\vartheta_2}{s} U^{\frac{s}{2}} \qquad (|\vartheta_2| \leqslant 1). \tag{5}$$

Considering (4) as a formal expansion of $U$ in a power series of $t$, we easily see that this series is majorized by the series

$$\sum_{k=2}^{\infty} \frac{1}{k!}\left(\frac{\beta_s^{\frac{1}{s}}|t|}{B_n}\right)^k,$$

and the expansion of $U^j$ in powers of $t$ is majorized by

$$\sum_{k=2j}^{\infty} \frac{1}{k!}\left(\frac{j\beta_s^{\frac{1}{s}}|t|}{B_n}\right)^k, \qquad 1 \leqslant j < \frac{s}{2}.$$

Therefore the series

$$\sum_{k=s}^{\infty} \frac{1}{k!}\left(\frac{j\beta_s^{\frac{1}{s}}|t|}{B_n}\right)^k$$

majorizes the sum of those terms in the expansion of $U$ which contain powers of $t$ of order $\geq s$. From this we obtain by simple calculations that for $|t| \leq T_{sn}$

$$\log f\left(\frac{t}{B_n}\right) = \sum_{k=2}^{s-1} \frac{\varkappa_k}{k!}\left(\frac{it}{B_n}\right)^k + \vartheta_3 \frac{s^s\beta_s}{s!}\left(\frac{t}{B_n}\right)^s$$

$$= -\frac{t^2}{2n} + \sum_{k=2}^{s-1} \frac{\lambda_k}{k!}\left(\frac{it}{\sqrt{n}}\right)^k + \vartheta_3\rho_s \frac{s^s}{s!}\left(\frac{t}{\sqrt{n}}\right)^s \quad (|\vartheta_3| \leqslant 1).$$

From this we find that

$$\log f_n(t) = -\frac{t^2}{2} + n\sum_{k=2}^{s-1} \frac{\lambda_k}{k!}\left(\frac{it}{\sqrt{n}}\right)^k + \vartheta_3\rho_s \frac{s^s}{s!} n \left(\frac{t}{\sqrt{n}}\right)^s.$$

Put

$$V = \log\left\{ e^{\frac{t^2}{2}} (f_n(tz))^{\frac{1}{z^2}} \right\} = \sum_{k=1}^{s-3} \frac{\lambda_{k+2}(it)^{k+2}}{(k+2)!} \left(\frac{z}{\sqrt{n}}\right)^k + \vartheta_3 \frac{s^s}{s!} \rho_s \left(\frac{z}{\sqrt{n}}\right)^{s-2}.$$

We expand $e^V$ in a power series of $z$, regarding $z$ as a real variable $|z| \leq 1$, and $t$ and $n$ as fixed. Then we obtain [cf. (25) of § 38]

$$e^V = e^{\frac{t^2}{2}} (f_n(tz))^{\frac{1}{z^2}} = 1 + \sum_{k=1}^{s-3} P_k(it) \left(\frac{z}{\sqrt{n}}\right)^k + R(z),$$

where $R(z) = 0(z^{s-2})$ as $z$ approaches 0.

The series for $V$ is majorized by the series

$$\frac{\rho_3}{3!}|t|^3 \frac{|z|}{\sqrt{n}} + \sum_{k=2}^{s-2} \frac{(k+2)^{k+2}}{(k+2)!} \rho_{k+2} |t|^{k+2} \left|\frac{z}{\sqrt{n}}\right|^k + \sum_{k=s-1}^{\infty} \frac{(\rho_s^{\frac{1}{s}} |t|)^{k+2}}{k!} \left|\frac{z}{\sqrt{n}}\right|^k.$$

We note that for $1 \leq k \leq s-2$,

$$\frac{(k+2)^{k+2}}{(k+2)!} \leqslant 3\frac{(k+2)^{k-1}}{(k-1)!} \leqslant 3\frac{s^{k-1}}{(k-1)!}$$

and

$$\rho_{k+2} \leqslant \rho_s^{\frac{k+2}{s}}.$$

Hence the series

$$3(\rho_s^{\frac{1}{s}}|t|)^3 \frac{|z|}{\sqrt{n}} \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{s\rho_s^{\frac{1}{s}}|tz|}{\sqrt{n}}\right)^k \tag{6}$$

*a fortiori* majorizes the series for $V$. From this it is clear that the series for $V^j$ is majorized by the series

$$3^j(\rho_s^{\frac{1}{s}}|t|)^{3j} \left|\frac{z}{\sqrt{n}}\right|^j \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{js\rho_s^{\frac{1}{s}}|tz|}{\sqrt{n}}\right)^k. \tag{7}$$

Now we estimate the remainder term $R(z)$. First of all, we have

$$e^V = \sum_{k=0}^{s-3} \frac{V^k}{k!} + \vartheta_4 V^{s-2} e^{|V|} \qquad \left(|\vartheta_4| \leqslant \frac{1}{(s-2)!}\right).$$

Furthermore,

$$\sum_{k=0}^{s-3} \frac{V^k}{k!} = 1 + \sum_{\nu=1}^{s-3} P_k(it)\left(\frac{z}{\sqrt{n}}\right)^k + \omega(z),$$

where $\omega(z)$ contains only powers of $z$ beginning with the $(s-2)$nd. Using (6) and (7), we find that

$$|\omega(z)| \leqslant \sum_{j=1}^{s-3} 3^j (\rho_s^{\frac{1}{s}} |t|)^{3j} \left|\frac{z}{\sqrt{n}}\right|^j \sum_{k=s-2-j}^{\infty} \frac{1}{k!} \left(\frac{js\rho_s^{\frac{1}{s}} |tz|}{\sqrt{n}}\right)^k.$$

Since $|z| \leq 1$, we have for $|t| \leq T_{sn}$,

$$\frac{js\rho_s^{\frac{1}{s}} |tz|}{\sqrt{n}} \leqslant \frac{j\rho_s^{\frac{1}{s}}}{8\rho_s^{\frac{3}{s}}} \leqslant \frac{j}{8} \leqslant \frac{s}{8} \tag{8}$$

and

$$\sum_{k=s-2-j}^{\infty} \frac{1}{k!} \left(\frac{js\rho_s^{\frac{1}{s}} |tz|}{\sqrt{n}}\right)^k \leqslant \left(\frac{js\rho_s^{\frac{1}{s}} |tz|}{\sqrt{n}}\right)^{s-2-j} \sum_{k=0}^{\infty} \frac{\left(\frac{s}{8}\right)^k}{(s+k-2-j)!}$$

$$\leqslant \left(\frac{s^2\rho_s^{\frac{1}{s}} |tz|}{\sqrt{n}}\right)^{s-2-j} e^{\frac{s}{8}}.$$

Therefore,

$$|\omega(z)| \leqslant \frac{3}{s^2} e^{\frac{s}{8}} \left(\frac{s^2\rho_s^{\frac{1}{s}} |tz|}{\sqrt{n}}\right)^{s-2} \sum_{j=1}^{s-3} (\rho_s^{\frac{1}{s}} |t|)^{2j}.$$

But for any geometrical progression with ratio $a > 0$,

$$\sum_{j=1}^{s-3} a^j \leqslant (s-3)(a + a^{s-2}),$$

so that

$$|\omega(z)| \leqslant \frac{3(s-3)}{s^2} e^{\frac{s}{8}} s^{(s-2)} \left(\frac{s\rho_s^{\frac{1}{s}} |t|}{\sqrt{n}}\right)^{s-2} \left((\rho_s^{\frac{1}{s}} |t|)^2 + (\rho_s^{\frac{1}{s}} |t|)^{2(s-2)}\right)$$

$$\leqslant \frac{3(s-3)}{s^2} e^{\frac{s}{8}} s^{(s-2)} \left(\frac{s\rho_s^{\frac{3}{s}}}{\sqrt{n}}\right)^{s-2} \left(\frac{|t|^s}{\rho_s^{\frac{2(s-3)}{s}}} + |t|^{3(s-2)}\right)$$

$$\leqslant \frac{c'(s)}{T_{sn}^{s-2}} (|t|^s + |t|^{3(s-2)}).$$

Now we estimate $V^{s-2}e^{|V|}$. According to (7) and (8)

$$|V|^{s-2}e^{|V|} \leqslant 3^{s-2}(\rho_s^{\frac{1}{s}}|t|)^{3(s-2)}\left(\frac{1}{\sqrt{n}}\right)^{s-2} e^{\frac{s}{8}+|V|}.$$

But, as follows from (6) and (8),

$$|V| \leqslant 3t^2 \frac{\rho_s^{\frac{3}{s}}|t|}{\sqrt{n}} e^{\frac{1}{8}} = \frac{3}{8s} t^2 e^{\frac{1}{8}} < \frac{t^2}{4}.$$

From the inequalities obtained we deduce that

$$|V|^{k-2}e^{|V|} \leqslant c''(s)\frac{|t|^{3(s-2)}}{T_{sn}^{s-2}} e^{\frac{t^2}{4}}.$$

Consequently,

$$|R(z)| \leqslant \frac{c_1(k)}{T_{kn}^{k-2}}\left[|t|^s + |t|^{3(s-2)} + |t|^{3(s-2)}e^{\frac{t^2}{4}}\right].$$

Since

$$\left|[f_n(tz)]^{\frac{1}{z^2}} - \left[1 + \sum_{\nu=1}^{s-3} P_\nu(it)\left(\frac{z}{\sqrt{n}}\right)^\nu\right]e^{-\frac{t^2}{2}}\right| = |R(z)|e^{-\frac{t^2}{2}},$$

we obtain the assertion of the theorem by taking $z = 1$.

Theorem 1(b) is proved similarly, except that in the proof it is necessary to use, instead of the expansion (4), the following expansion in the neighborhood of $t = 0$:

$$U = \sum_{k=2}^{s} \frac{a_k}{k!}\left(\frac{it}{B_n}\right)^k + o\left(\frac{t}{B_n}\right)^k.$$

We shall not enter into the details.

## § 42. Improvement of Lyapunov's Theorem for Nonlattice Distributions

The basic object of this section is to prove Theorem 2; for this purpose we need the following auxiliary proposition (Esseen [26]).

THEOREM 1. *If the distribution function $F(x)$ is nonlattice, then whatever the number $\omega > 0$ may be, there exists a function $\lambda(n)$ such that* $\lim_{n\to\infty} \lambda(n) = \infty$ *and*

$$I = \int_{\omega}^{\lambda(n)} \frac{|f^n(t)|}{t}\,dt = o\left(\frac{1}{\sqrt{n}}\right). \quad (1)$$

*Proof.* If the function $F(x)$ is such that *

$$\overline{\lim_{|t|\to\infty}} |f(t)| < 1,$$

then the theorem to be proved becomes trivial.

In fact, from Theorem 5 of § 14 and the condition (C) we deduce that to every $\epsilon > 0$ there corresponds a $c(\epsilon) > 0$ such that $|f(t)| \leq e^{-c(\epsilon)} < 1$ for $|t| \geq \epsilon$.

Hence, in particular, there exists $c > 0$ such that $|f| \leq e^{-c}$ for $|t| \geq \omega$. Putting $\lambda(n) = n$, we find that

$$I \leqslant \int_{\omega}^{n} \frac{e^{-cn}}{t}\, dt = e^{-cn} \log \frac{n}{\omega} = o\left(\frac{1}{\sqrt{n}}\right).$$

Now let

$$\overline{\lim_{|t|\to\infty}} |f(t)| = 1.$$

Since by hypothesis the distribution considered is nonlattice, the equation $|f(t_0)| = 1$ cannot hold for any $t_0 \neq 0$ (see Theorem 5 of § 14). Hence it is possible to define a function $a(t)$ for $t \geq \omega$ by the following equation:

$$1 - \frac{1}{a(t)} = \max_{\omega \leqslant \tau < t} |f(\tau)|.$$

Obviously the function $a(t)$ is continuous, nondecreasing, and by virtue of the condition $\overline{\lim\limits_{t\to\infty}} |f(t)| = 1$ satisfies the relation

$$\lim_{t\to\infty} a(t) = \infty.$$

From the definition of the function $a(t)$ it is evident that

$$I = \int_{\omega}^{\lambda(n)} \frac{|f(t)|^n}{t}\, dt \leqslant \int_{\omega}^{\lambda(n)} \frac{\left(1 - \frac{1}{a(t)}\right)^n}{t}\, dt$$

for every $\lambda(n) \geq \omega$. If $a(n) \leq \sqrt{n}$, we set $\lambda(n) = n$. Then we obtain

$$I \leqslant \int_{\omega}^{n} \frac{1}{t}\left(1 - \frac{1}{\sqrt{n}}\right)^n dt \leqslant e^{-\frac{\sqrt{n}}{2}} \log \frac{n}{\omega} = o\left(\frac{1}{\sqrt{n}}\right).$$

---

* H. Cramér calls this *Condition (C)*; the results of § 45 can be obtained under the assumption that Condition (C) is satisfied.

If, however, $a(n) > \sqrt{n}$, then

$$I \leqslant \int_{\omega}^{\lambda(n)} \frac{1}{t}\left(1 - \frac{1}{a(\lambda(n))}\right)^n dt = \left(1 - \frac{1}{a(\lambda(n))}\right)^n \log \frac{\lambda(n)}{\omega}.$$

Let $t(a)$ be the inverse function of $a(t)$. Evidently $\lim\limits_{a\to\infty} t(a) = \infty$. We now set $\lambda(n) = t(\sqrt{n})$. Then $a(\lambda(n)) = a(t(\sqrt{n})) = \sqrt{n}$. Furthermore, since $a(n) > \sqrt{n}$, $t(\sqrt{n}) < n$ and consequently $\lambda(n) < n$. Therefore, in this case also,

$$I \leqslant \left(1 - \frac{1}{\sqrt{n}}\right)^n \log \frac{n}{\omega} = o\left(\frac{1}{\sqrt{n}}\right).$$

Q.E.D.

The theorem formulated below was proved by H. Cramér [21] under the assumption that Condition (C) is satisfied, and by G. Esseen [26] in the form presented here.

THEOREM 2. *If the independent random variables $\xi_1, \xi_2, \ldots, \xi_n$ are identically distributed, nonlattice, and have finite third moments, then*

$$F_n(x) - \Phi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \cdot \frac{Q_1(x)}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right) \tag{2}$$

*uniformly in $x$. Here* $Q_1(x) = \frac{\lambda_3}{6}(1 - x^2) = \frac{\alpha_3}{6\sigma^3}(1 - x^2)$ [cf. (1) and (24) of § 38].

*Proof.* Putting $s = 3$ in Theorem 1(b) of § 41, we find that

$$\left| f_n(t) - e^{-\frac{t^2}{2}} - \frac{P_1(it)}{\sqrt{n}} e^{-\frac{t^2}{2}} \right| \leqslant \frac{\delta(n)}{\sqrt{n}} (|t|^3 + |t|^6) e^{-\frac{t^2}{4}}. \tag{3}$$

The characteristic function * of the function

$$P_1(-\Phi) = -\frac{\alpha_3}{6\sigma^3} \Phi^{(3)}(x) = \frac{\alpha_3}{6\sigma^3 \sqrt{2\pi}} (1 - x^2) e^{-\frac{x^2}{2}}$$

is equal to (see § 38)

$$\frac{\alpha_3}{6\sigma^3} (it)^3 e^{-\frac{t^2}{2}} = P_1(it) e^{-\frac{t^2}{2}}.$$

---

* *Translator's note.* The characteristic function of a function $F(x)$ of bounded variation, not necessarily a distribution function, is its Fourier-Stieltjes transform $\int e^{itx} dF(x)$.

Use Theorem 1 of § 39, and put there

$$F(x) = F_n(x), \quad G(x) = \Phi(x) + \frac{1}{\sqrt{n}} P_1(-\Phi),$$

$$A = \max |G'(x)| < +\infty, \quad T = \lambda(n)\sqrt{n}$$

[$\lambda(n)$ is defined as in Theorem 1 taking $\omega = 1/24\rho_3$].

Without loss of generality we may suppose that $T \geq T_{3n}$. (In fact, this inequality is satisfied for all sufficiently large $n$.)

We estimate the quantity

$$\varepsilon = \int_{-T}^{T} \left| \frac{f_n(t) - g(t)}{t} \right| dt = \int_{-T}^{-T_{3n}} + \int_{-T_{3n}}^{T_{3n}} + \int_{T_{3n}}^{T}.$$

According to (3),

$$\int_{-T_{3n}}^{T_{3n}} \left| \frac{f_n(t) - g(t)}{t} \right| dt \leqslant \frac{\delta(n)}{\sqrt{n}} \int (t^2 + |t|^5) e^{-\frac{t^2}{4}} dt = o\left(\frac{1}{\sqrt{n}}\right).$$

Furthermore,

$$\int_{T_{3n}}^{T} \left| \frac{f_n(t) - g(t)}{t} \right| dt \leqslant \int_{T_{3n}}^{T} \left| f\left(\frac{t}{B_n}\right) \right|^n \frac{dt}{t} + \int_{T_{3n}}^{T} |g(t)| \frac{dt}{t}.$$

But

$$\int_{T_{3n}}^{T} |g(t)| \frac{dt}{t} \leqslant \frac{1}{\sqrt{2\pi}\, T_{3n}} \int_{T_{3n}}^{T} e^{-\frac{t^2}{2}} \left(1 + \frac{\alpha_3 |t|^3}{6\sigma^3 \sqrt{n}}\right) dt = o\left(\frac{1}{\sqrt{n}}\right)$$

and by the preceding theorem

$$\int_{T_{3n}}^{T} \left| f\left(\frac{t}{B_n}\right) \right|^n \frac{dt}{t} = \int_{\omega}^{\lambda(n)} |f(t)|^n \frac{dt}{t} = o\left(\frac{1}{\sqrt{n}}\right).$$

The integral $\int_{-T}^{-T_{3n}}$ is estimated similarly. Thus $\epsilon = o(1/\sqrt{n})$. An application of Theorem 1 of § 39 leads to the inequality

$$\left| F_n(x) - \Phi(x) - \frac{P_1(-\Phi)}{\sqrt{n}} \right| \leqslant \frac{a}{2\pi} \varepsilon + \frac{c(a)A}{\lambda(n)\sqrt{n}} = o\left(\frac{1}{\sqrt{n}}\right),$$

which proves the theorem.

## § 43. Deviation from the Limit Law in the Case of a Lattice Distribution

We shall preface the detailed exposition of the following results with a brief intuitive argument which should clarify the reasons why Theorem 2 of § 42 does not hold for lattice distributions. Suppose that the random variables $\xi_k$ can take only two values $+1$ and $-1$, each with probability $\frac{1}{2}$. If $n$ is an even number, then the functions $F_n(x)$ are discontinuous, with jumps at the points $x_v = v/\sqrt{n}$, $(v = 0, \pm 2, \pm 4, \ldots, \pm n)$. According to the local theorem of de Moivre-Laplace, at each discontinuity point $x$ the function $F_n(x)$ has a jump asymptotically equal to $\frac{2}{\sqrt{2\pi n}} e^{-\frac{x_v^2}{2}}$ (this assertion is a particular case of the theorem to be proved in § 49).

Consider in more detail the relative behavior of the functions $F_n(x)$ and $\Phi(x)$ in the neighborhood of a discontinuity point, say the point $x = 0$. In the interval $(-(1/\sqrt{n}), (1/\sqrt{n}))$ the function $\Phi(x)$ behaves like the function

$$\frac{x}{\sqrt{2\pi}} + \frac{1}{2}$$

up to infinitesimals of higher order. Introduce the function

$$S(x) = [x] - x + \frac{1}{2}.$$

It is easily seen that up to infinitesimals of higher order than $1/\sqrt{n}$ the equation

$$F_n(x) - \Phi(x) = \frac{2}{\sqrt{2\pi n}} S\left(\frac{x\sqrt{n}}{2}\right)$$

holds in the interval $(-(1/\sqrt{n}), (1/\sqrt{n}))$.

If we wish to write down analogous asymptotic equations for other values of $x$, then we must take into account the change of the slope of the curve $y = \Phi(x)$. Thus we are led to the consideration of the difference

$$F_n(x) - \Phi(x) - D_n^{(0)}(x),$$

where

$$D_n^{(0)}(x) = \frac{2}{\sqrt{2\pi n}} S\left(\frac{x\sqrt{n}}{2}\right) e^{-\frac{x^2}{2}}. \tag{1}$$

We now consider the general case of a lattice distribution. Let the possible values of the random variable $\xi_k$ be $x_v = a + vh$ ($v = 0, \pm 1, \pm 2, \ldots$) and the span $h$ of the distribution be maximum. We put all lattice distributions with the maximum span $h$ into one class and call it the class $L_h$. If

$$\mathsf{P}\{\xi_k = a + \nu h\} = p_\nu,$$

then by the tacit assumption $\mathbf{M}\xi_k = 0$

$$\bar{p} = \sum_{\nu=-\infty}^{\infty} \nu p_\nu = -\frac{a}{h}.$$

The possible values of the sums $\zeta_n$ will have the form

$$\frac{h}{\sigma\sqrt{n}}(\nu - n\bar{p}). \tag{2}$$

Put

$$a_n = \frac{hn\bar{p}}{\sigma\sqrt{n}}$$

and consider the function

$$S_1(x) = \frac{h}{\sigma} S\left(\frac{(x + a_n)\sigma\sqrt{n}}{h}\right). \tag{3}$$

The object of this section is to prove the following theorem (Esseen [26]).

THEOREM 1. *If $\xi_1, \xi_2, \ldots, \xi_n$ are independent, identically distributed random variables having finite third moments and belonging to the class $L_h$, then*

$$F_n(x) - \Phi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\left(\frac{Q_1(x)}{\sqrt{n}} + \frac{S_1(x)}{\sqrt{n}}\right) + o\left(\frac{1}{\sqrt{n}}\right) \tag{4}$$

*uniformly in $x$.*

*Proof.* First of all we calculate the characteristic function of

$$D_n(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi n}} S_1(x):$$

$$d_n(t) = \int e^{itx} dD_n(x) = -it \int e^{itx} D_n(x)\, dx.$$

For this purpose we note that $S(x)$ is periodic with period one and write the function $S[(x + a_n/h)\sigma\sqrt{n}]$ in the form of a Fourier series. By the usual methods it is easily found that

$$S(x) = \sum_{\nu=1}^{\infty} \frac{1}{\nu\pi} \sin 2\pi\nu x$$

and that consequently

$$S\left(\frac{x + a_n}{h}\sigma\sqrt{n}\right) = \sum_{\nu=1}^{\infty} \frac{1}{\nu\pi} \sin(\tau\sigma\sqrt{n}\,\nu(x + a_n)),$$

where

$$\tau = \frac{2\pi}{h}.$$

Therefore

$$d_n(t) = -\frac{2}{\tau\sigma\sqrt{2\pi n}}\sum_{\nu=1}^{\infty}\frac{it}{\nu}\int e^{itx-\frac{x^2}{2}}\sin(\tau\sigma\sqrt{n}\,\nu(x+a_n))\,dx$$

$$= -\frac{t}{\tau\sigma\sqrt{2\pi n}}\sum_{\nu=-\infty}^{\infty}{}'\frac{1}{\nu}\int e^{itx-\frac{x^2}{2}+i\nu\tau\sigma\sqrt{n}(x+a_n)}\,dx,$$

where the summation is extended to all integral values $v \neq 0$. An easy calculation leads to the equation

$$d_n(t) = -\frac{t}{\tau\sigma\sqrt{n}}\sum_{\nu=-\infty}^{\infty}{}'\frac{e^{i\tau\nu\sigma\sqrt{n}\,a_n}}{\nu}e^{-\frac{1}{2}(t+\tau\sigma\nu\sqrt{n})^2}.$$

We now apply Theorem 2 of § 39, putting there

$$F(x) = F_n(x),\quad G(x) = \Phi(x) + \frac{P_1(-\Phi)}{\sqrt{n}} + D_n(x),$$

$$l = \frac{h}{\sigma\sqrt{n}},\quad A = \max_{x\neq x_\nu}|G'(x)| < +\infty\ (\nu = 0, \pm 1, \pm 2, \ldots),$$

$$T = n > T_{3n} = \frac{\sqrt{n}}{24\rho_3}.$$

Then, whatever $c_2(k)$ may be, for sufficiently large $n$

$$Tl = \frac{h\sqrt{n}}{\sigma} \geqslant c_2(k).$$

We now estimate the integral

$$\varepsilon = \int_{-T}^{T}\left|\frac{f_n(t)-g(t)}{t}\right|dt.$$

For this purpose we split it into three parts:

$$\varepsilon_1 = \int_{-T}^{-\frac{\tau}{2}\sigma\sqrt{n}},\quad \varepsilon_2 = \int_{-\frac{\tau}{2}\sigma\sqrt{n}}^{\frac{\tau}{2}\sigma\sqrt{n}},\quad \varepsilon_3 = \int_{\frac{\tau}{2}\sigma\sqrt{n}}^{T}\left|\frac{f_n(t)-g(t)}{t}\right|dt.$$

We may suppose that $T_{3n} < \frac{1}{2}\tau\sigma\sqrt{n}$ (otherwise the estimates will only be simplified). The period of the function $|f(t)|$ is $\tau$; consequently, according to Remark 2 after Theorem 5 of § 14, it is possible to find a number $c_1 > 0$ (not dependent on $n$ of the condition $T_{3n}/\sigma\sqrt{n} = 1/24\rho_3\sigma$), such that $|f(t)| < e^{-c_1}$ for $T_{3n}/\sigma\sqrt{n} \le |t| \le \tau/2$. Then in the interval

$$T_{3n} \leqslant |t| \leqslant \frac{1}{2}\tau\sigma\sqrt{n},$$

$$\left|f\left(\frac{t}{\sigma\sqrt{n}}\right)\right|^n < e^{-c_1 n}.$$

It is also obvious that in this interval for some $c_2 > 0$

$$|g(t)| \leqslant e^{-c_2 n}.$$

Therefore

$$\varepsilon_2 \leqslant \int_{-T_{3n}}^{T_{3n}} \left| \frac{f_n(t)-g(t)}{t} \right| dt + 4 \int_{T_{3n}}^{\frac{1}{2}\tau\sigma\sqrt{n}} \frac{e^{-cn}}{t}\, dt.$$

By the definition of $g(t)$ and Theorem 1(b) of § 41

$$\int_{-T_{3n}}^{T_{3n}} \left| \frac{f_n(t)-g(t)}{t} \right| dt = o\left(\frac{1}{\sqrt{n}}\right) + \int_{-T_{3n}}^{T_{3n}} \left| \frac{d_n(t)}{t} \right| dt.$$

But for $|t| \leq T_{3n}$ and sufficiently large $n$,

$$\left| \frac{d_n(t)}{t} \right| \leqslant \frac{1}{\sigma\tau\sqrt{n}} \sum_{\nu=-\infty}^{\infty}{}' \frac{1}{|\nu|} e^{-\frac{\tau^2\sigma^2 n\nu^2}{2}} e^{-\frac{t^2}{2}} = o\left(\frac{1}{\sqrt{n}}\right) e^{-\frac{t^2}{2}}.$$

Therefore, finally,

$$\varepsilon_2 = o\left(\frac{1}{\sqrt{n}}\right).$$

We now turn to the estimation of $\epsilon_3$:

$$\varepsilon_3 = \int_{\frac{\tau}{2}\sigma\sqrt{n}}^{n} \left| \frac{f_n(t)-g(t)}{t} \right| dt = O\left(\frac{1}{n}\right) + \int_{\frac{\tau}{2}\sigma\sqrt{n}}^{n} \left| \frac{f_n(t)-d_n(t)}{t} \right| dt$$

$$= O\left(\frac{1}{n}\right) + \int_{\frac{\tau}{2}}^{\frac{\sqrt{n}}{\sigma}} \left| \frac{f^n(t)-d_n(\sigma\sqrt{n}\,t)}{t} \right| dt$$

$$= O\left(\frac{1}{n}\right) + \sum_{k=1}^{r} \int_{\frac{2k-1}{2}\tau}^{\frac{2k+1}{2}\tau} + \int_{\frac{2r+1}{2}\tau}^{\frac{\sqrt{n}}{\sigma}} \left| \frac{f^n(t)-d_n(\sigma\sqrt{n}\,t)}{t} \right| dt,$$

where

$$r = \left[\frac{\sqrt{n}}{\sigma\tau} - \frac{1}{2}\right].$$

Put

$$I_k = \int_{\frac{2k-1}{2}\tau}^{\frac{2k+1}{2}\tau} \left| \frac{f^n(t)-d_n(\sigma\sqrt{n}\,t)}{t} \right| dt$$

and make the change of variables $t = z + k\tau$. Here we recall that

$$f(t) = \sum_{\nu=-\infty}^{+\infty} e^{i(\nu-\bar{p})th} p_\nu,$$

and consequently

$$f(z+k\tau)^n = e^{-ik\tau pnh} f^n(z) = e^{-ik\tau a_n \sigma \sqrt{n}} f^n(z).$$

Therefore

$$I_k = \int_{-\frac{\tau}{2}}^{+\frac{\tau}{2}} \left| \frac{e^{-ik\tau a_n \sigma\sqrt{n}} f^n(z) + \frac{z+k\tau}{\tau} \sum_\nu' \frac{1}{\nu} e^{i\nu\tau a_n \sigma \sqrt{n}} e^{-\frac{n\sigma^2}{2}(z+k\tau+\nu\tau)^2}}{z+k\tau} \right| dz.$$

But

$$\sum_\nu' \frac{1}{\nu} e^{i\nu\tau a_n \sigma\sqrt{n}} e^{-\frac{n\sigma^2}{2}(z+k\tau+\nu\tau)^2}$$

$$= -\frac{e^{-ik\tau a_n \sigma\sqrt{n}}}{k} e^{-\frac{n\sigma^2}{2}z^2} + O\left(e^{-\frac{n\sigma^2}{2}\frac{\tau^2}{4}}\right),$$

uniformly in $z$ ($|z| \leq \tau/2$); hence

$$I_k = \int_{-\frac{\tau}{2}}^{+\frac{\tau}{2}} \left| \frac{(f(z))^n - e^{-\frac{1}{2}\sigma^2 n z^2} - \frac{z}{k\tau} e^{-\frac{1}{2}\sigma^2 n z^2}}{z+k\tau} \right| dz + O\left(e^{-\frac{n\sigma^2\tau^2}{8}}\right).$$

Now, on the one hand, by Theorem 2 of § 40

$$\int_{-\frac{\tau}{2}}^{+\frac{\tau}{2}} \left| \frac{(f(z))^n - e^{-\frac{1}{2}\sigma^2 n z^2}}{z+k\tau} \right| dz$$

$$\leqslant \frac{2}{k\tau} \int_{-\frac{\tau}{2}\sigma\sqrt{n}}^{+\frac{\tau}{2}\sigma\sqrt{n}} \left| f_n(z) - e^{-\frac{1}{2}z^2} \right| \frac{dz}{\sigma\sqrt{n}} = O\left(\frac{1}{kn}\right).$$

On the other hand,

$$\frac{1}{k\tau} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \frac{|t|}{|t|+k\tau} e^{-\frac{1}{2}\sigma^2 n t^2} dt = O\left(\frac{1}{k^2 n}\right).$$

Therefore

$$I_k = O\left(\frac{1}{kn}\right)$$

and since $r = O(\sqrt{n})$,

$$\varepsilon_3 = O\left(\frac{1}{n}\sum_{k=1}^{r}\frac{1}{k}\right) = O\left(\frac{\log n}{n}\right).$$

In exactly the same way,

$$\varepsilon_1 = O\left(\frac{\log n}{n}\right).$$

Therefore, finally,

$$\varepsilon = o\left(\frac{1}{\sqrt{n}}\right).$$

An application of Theorem 2 of § 39 leads to the inequality

$$|F_n(x) - G(x)| \leqslant \frac{a}{2\pi}\, o\left(\frac{1}{\sqrt{n}}\right) + c_1(a)\frac{A}{n} = o\left(\frac{1}{\sqrt{n}}\right),$$

which proves our theorem.

## § 44. The Extremal Character of the Bernoulli Case

The results of the last two sections enable us to obtain some extremal properties of lattice distributions and to clarify the special role of the Bernoulli distribution (Esseen [26]).

Theorem 1. *If the random variables $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ are independent, identically distributed, and have finite third moments, then*

$$\lim_{n\to\infty}\ \max_{-\infty<x<+\infty} \sqrt{n}\left|F_n(x) - \Phi(x) - \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\frac{Q_1(x)}{\sqrt{n}}\right| = \omega(h),$$

*where $\omega(h)$ is equal to 0 if $F(x)$ is a nonlattice distribution, and equal to $\dfrac{h}{2\sigma\sqrt{2\pi}}$ if $F(x)$ is a lattice distribution with the maximum span $h$.*

*Proof.* Indeed, for nonlattice distributions, as it follows from Theorem 2 of § 42,

$$R_n(x) = \sqrt{n}\left|F_n(x) - \Phi(x) - \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\cdot\frac{Q_1(x)}{\sqrt{n}}\right| = o(1).$$

From the theorem in § 43 we find that for lattice distributions under the conditions of the present theorem

$$R_n = \frac{h}{\sigma\sqrt{2\pi}}\, S\left(\frac{(x+a_n)\sigma\sqrt{n}}{h}\right)e^{-\frac{x^2}{2}} + o(1). \tag{1}$$

From the definition of the function $S(x)$ and its periodicity with period $h/\sigma\sqrt{n}$, we deduce that

$$\max_{-\infty<x<\infty} R_n = \frac{h}{2\sigma\sqrt{2\pi}}\, e^{-\frac{\theta_n^2}{2}} + o(1),\quad 0 \leqslant \theta_n < \frac{h}{\sigma\sqrt{n}}. \tag{1'}$$

Since $\theta_n \to 0$ as $n \to \infty$,

$$\lim_{n\to\infty} \max_{-\infty<x<\infty} R_n = \frac{h}{2\sigma\sqrt{2\pi}},$$

proving the theorem.

Thus, the convergence of the distribution functions of normalized sums to a limit law is in a certain sense worst for lattice distributions. The following theorem in a way supplements Theorem 1 in the case of symmetrical distributions.

THEOREM 2. *If the random variables $\xi_1, \xi_2, \ldots, \xi_n$ are symmetrically distributed and satisfy the conditions of the preceding theorem, and if their distribution function is continuous at the point $x = 0$, then*

$$\lim_{n\to\infty} \max_{-\infty<x<\infty} \sqrt{n}\,|F_n(x) - \Phi(x)| \leqslant \frac{1}{\sqrt{2\pi}}.$$

*The last inequality becomes an equality if and only if*

$$F(x) = \begin{cases} 0 & \text{for } x \leqslant -\dfrac{h}{2}, \\ \dfrac{1}{2} & \text{for } |x| < \dfrac{h}{2}, \\ 1 & \text{for } x > \dfrac{h}{2} \end{cases}$$

(Bernoulli scheme).

*Proof.* As the preceding theorem shows, we may confine ourselves to the consideration of lattice distributions. Therefore, we have to find the maximum value of the quantity

$$\frac{h}{2\sigma\sqrt{2\pi}}.$$

Two cases may occur: 1) the variable $\xi_k$ takes values of the form $vh$ ($v = \pm 1, \pm 2, \ldots$) with probabilities $p_v$; 2) $\xi_k$ takes values of the form $(v - \frac{1}{2})h$ ($v = 0, \pm 1, \pm 2, \ldots$) with probabilities $p_v$; in either case $\Sigma p_v = 1$. In the first case,

$$\sigma^2 = 2\sum_{\nu=1}^{\infty} \nu^2 h^2 p_\nu \geqslant 2h^2 \sum_{\nu=1}^{\infty} p_\nu = h^2.$$

In the second case,

$$\sigma^2 = \sum_{\nu=-\infty}^{\infty} \left(\nu - \frac{1}{2}\right)^2 h^2 p_\nu \geqslant h^2 \sum_{\nu=-\infty}^{\infty} \left(\frac{1}{2}\right)^2 p_\nu = \frac{h^2}{4}.$$

In the first case, equality is unattainable:

$$\frac{h}{\sigma} < 1.$$

In the second case,

$$\frac{h}{\sigma} \leqslant 2;$$

moreover equality is attained if and only if all $p_\nu = 0$ for $\nu \neq 0$ or 1. Therefore we have always

$$\frac{h}{2\sigma\sqrt{2\pi}} \leqslant \frac{1}{\sqrt{2\pi}},$$

and equality is attained only for the symmetrical scheme of Bernoulli.

Both conditions of Theorem 2 (symmetry and the continuity of $F(x)$ at $x = 0$) are essential; the rejection of either one of these conditions would cause the expression

$$\lim_{n\to\infty} \max_{-\infty<x<+\infty} \sqrt{n}\,|F_n(x) - \Phi(x)|$$

to become unbounded.

Indeed, suppose that $\xi_k$ can take the value 0 with positive probability. The example of the symmetrical variables taking only the values $-1, +1, 0$ with the corresponding probabilities $p^2/2$, $p^2/2$, $1 - p^2$ shows that the expression

$$\frac{h}{2\sigma\sqrt{2\pi}} = \frac{1}{2p\sqrt{2\pi}} \tag{2}$$

becomes as large as we wish for sufficiently small $p$.

Similarly, if we discard the assumption of symmetry of the law $F(x)$, we can also easily see by a simple example that (2) becomes unbounded for laws of the class $L_h$. Indeed, let $\xi_k$ take only three values $a - h$, $a$, $a + kh$ with the corresponding probabilities $p_1$, $p_0$, $p_2$. We choose $a$, $p_1$, $p_2$, and $k$ so that

$$\mathsf{M}\xi_k = a + h(-p_1 + kp_2) = 0,$$

$$\mathsf{M}\xi_k^3 = a^3 + 3a\sigma^2 + h^3(-p_1 + k^3p_2) = 0.$$

Here $p_1$ and $p_2$ can still vary quite freely. Choose $p_1$ and $p_2$ so small that $k^2p_2$ is sufficiently small; then the expression

$$\frac{h}{2\sigma\sqrt{2\pi}} = \frac{1}{2\sqrt{2\pi(p_1(1-p_1) + k^2p_2(1-p_2) + 2kp_1p_2)}}$$

becomes as large as we wish.

## § 45. Improvement of Lyapunov's Theorem with Higher Moments for the Continuous Case

Now we shall formulate precisely and prove the theorem, discussed in § 38, concerning the expansion of the function $F_n(x)$ in a series of polynomials. In this connection it is necessary to impose on the summands $\xi_k$ stronger conditions than those imposed in Theorem 2 of § 42.

Theorem. *If the independent random variables $\xi_1, \xi_2, \ldots, \xi_n$ are identically distributed and have finite absolute moments $\beta_s$ of the sth order ($s \geq 3$) and if Condition (C),*

$$\overline{\lim_{|t|\to\infty}} |f(t)| < 1,$$

*is satisfied, then*

$$F_n(x) - \Phi(x)$$
$$= \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\left(\frac{Q_1(x)}{n^{\frac{1}{2}}} + \frac{Q_2(x)}{n} + \ldots + \frac{Q_{s-2}(x)}{n^{\frac{s-2}{2}}}\right) + o\left(\frac{1}{n^{\frac{s-2}{2}}}\right) \tag{1}$$

*uniformly in $x$.*

*Proof.* The proof of this proposition is based on the application of Theorems 1 of § 39 and 1(b) of § 42. We put in Theorem 1 of § 39

$$F(x) = F_n(x),$$

$$G(x) = \Phi(x) + \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\left(\frac{Q_1(x)}{n^{\frac{1}{2}}} + \frac{Q_2(x)}{n} + \ldots + \frac{Q_{s-2}(x)}{n^{\frac{s-2}{2}}}\right)$$
$$= \Phi(x) + \sum_{\nu=1}^{s-2} \frac{P_k(-\Phi)}{n^{\frac{\nu}{2}}},$$

$$A = \max_{-\infty < x < \infty} |G'(x)| < +\infty, \quad T = n^s.$$

The characteristic function of the function $G(x)$, as it is easily deduced from the definition of $P(-\Phi)$ and (13) of § 38, is equal to

$$e^{-\frac{t^2}{2}}\left(1 + \sum_{k=1}^{s-2} P_k(it)\left(\frac{1}{\sqrt{n}}\right)^k\right). \tag{2}$$

Without loss of generality, we may suppose that $T > T_{sn}$, where $T_{sn}$ is understood to be a quantity introduced in the formulation of Theorem 1 of § 41. We estimate the integral

$$\varepsilon = \int_{-T}^{T} \left|\frac{f_n(t) - g(t)}{t}\right| dt,$$

where, according to (2),

$$g(t)=e^{-\frac{t^2}{2}}\left[1+\sum_{k=1}^{s-2}\frac{P_k(it)}{n^{\frac{k}{2}}}\right].$$

By Theorem 1(b) of § 41 we easily find that

$$\int_{-T_{sn}}^{T_{sn}}\left|\frac{f_n(t)-g(t)}{t}\right|dt=o\left(\frac{1}{n^{\frac{s-2}{2}}}\right).$$

Furthermore, from Condition (C) it follows by Theorem 5 of § 14 that the distribution of the variable $\xi_k$ is not a lattice distribution; hence there exists a $c > 0$ such that for $|t| > \frac{1}{8s\sigma\rho_s^{3/s}}$

$$|f(t)|<e^{-c}.$$

Then for $|t| > T_{sn}$

$$|f_n(t)|=\left|f^n\left(\frac{t}{B_n}\right)\right|<e^{-cn}.$$

But

$$\int_{T_{sn}<|t|\leqslant T}\left|\frac{f_n(t)-g(t)}{t}\right|dt\leqslant\int_{T_{sn}<|t|\leqslant T}\frac{|f_n(t)|}{t}dt+\int_{T_{sn}<|t|\leqslant T}\frac{|g(t)|}{t}dt$$
$$<2e^{-cn}\log\frac{T}{T_{sn}}+o\left(\frac{1}{n^{\frac{s-2}{2}}}\right)=o\left(\frac{1}{n^{\frac{s-2}{2}}}\right).$$

Thus

$$\varepsilon=o\left(\frac{1}{n^{\frac{s-2}{2}}}\right).$$

Now according to Theorem 1 of § 39,

$$\left|F_n(x)-\Phi(x)-\sum_{k=1}^{s-2}\frac{P_k(-\Phi)}{n^{\frac{k}{2}}}\right|\leqslant\frac{a}{2\pi}\varepsilon+c(a)\frac{A}{T}=o\left(\frac{1}{n^{\frac{s-2}{2}}}\right),\quad(3)$$

which is what was to be proved.

In the preceding section we have seen that even in the case $s = 3$ a more complicated expansion of the function $F_n(x)$ holds for lattice distributions. In case $s > 3$ the last theorem cannot be extended in a considerable number of other cases. In case Condition (C) is not satisfied, i.e., if

$$\overline{\lim_{|t|\to\infty}}|f(t)|=1,\quad(4)$$

the order of the remainder term in the expansion of $F_n(x)$ turns out to depend on the arithmetical nature of the set of possible values of the random variable $\xi_k$. We remark in this connection that (4) can be satisfied for a nonlattice distribution only if all the variation of the $F(x)$ is concentrated in a set of measure zero (Cramér [21], Theorem 7). For example, if $\xi_k$ takes only the values $\pm 1$ and $\pm\sqrt{3}$ each with probability $\frac{1}{4}$, then its distribution is not a lattice distribution. Its characteristic function

$$f(t) = \frac{1}{2}(\cos t + \cos t\sqrt{3})$$

(as an almost periodic function) satisfies equation (4). Simple calculations * show that for even $n$ the function $F_n(x)$ has a jump at $x = 0$, asympototically equal to $2/\pi n$. This obviously means that even though all the moments of $F(x)$ are finite in our example, it is impossible to write the expansion

$$F_n(x) - \Phi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\left(\frac{Q_1(x)}{n^{\frac{1}{2}}} + \frac{Q_2(x)}{n}\right) + o\left(\frac{1}{n}\right).$$

We thus see that in the case of discrete distributions it is necessary to supplement the expansion (1) with discontinuous terms.

## § 46. Limit Theorem for Densities

If the random variables to be summed are continuous (i.e., if they have probability densities), it is natural to seek the conditions under which not only the distribution functions of the sums converge to a limit, but also the densities of the probability distributions of the sums converge to the density of the limit distribution. We see that the second requirement is not a consequence of the first by considering the following example.

---

* Here is a sketch of these calculations. It is evident that

$$[f(t)]^{2r} = \frac{1}{4^{2r}}\sum_{k=0}^{2r} C_{2r}^k (e^{it} + e^{-it})^k (e^{it\sqrt{3}} + e^{-it\sqrt{3}})^{2r-k}.$$

The magnitude of the jump of $F_n(x)$ at $x = 0$ is equal to the coefficient of $e^0$. The summands containing $e^0$ are obtained only by multiplying together the middle terms of the expansions of $(e^{it} + e^{-it})^s$ and $(e^{it\sqrt{3}} + e^{-it\sqrt{3}})^{2r-s}$ for even $s$.

Therefore, the required jump is equal to

$$p_0 = \frac{1}{4^{2r}}\sum_{k=0}^{r} C_{2r}^{2k}\cdot C_{2k}^k \cdot C_{2(r-k)}^{r-k} = \frac{1}{4^{2r}} C_{2r}^r \sum_{k=0}^{r} (C_r^k)^2 = \frac{(C_{2r}^r)^2}{4^{2r}}.$$

An application of Stirling's formula leads to the asymptotic equation

$$p_0 \sim \frac{1}{\pi r} = \frac{2}{\pi n}.$$

Let the distribution function $F(x)$ be defined by means of the equation

$$F(x) = \int_{-\infty}^{x} p(z)\,dz,$$

where

$$p(x) = \begin{cases} 0 & \text{for } |x| \geqslant \dfrac{1}{e}, \\ \dfrac{1}{2|x|\log^2|x|} & \text{for } |x| < \dfrac{1}{e}. \end{cases}$$

Since the random variable $\xi$ with such a distribution is bounded, the function $F(x)$ belongs to the domain of normal attraction of the normal law. In other words, if $\xi_1, \xi_2, \ldots, \xi_n$ are independent random variables having $F(x)$ as their common distribution, and

$$\sigma^2 = \mathsf{D}^2\xi_n = \mathsf{M}\xi_n^2 = \int_0^{\frac{1}{e}} \frac{x}{\log^2 x}\,dx,$$

then as $n \to \infty$

$$\mathsf{P}\left\{\frac{\xi_1 + \xi_2 + \ldots + \xi_n}{\sigma\sqrt{n}} < x\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}}\,dz.$$

We shall show that the probability densities of the sums

$$\frac{1}{\sigma\sqrt{n}}(\xi_1 + \xi_2 + \ldots + \xi_n)$$

do not converge to the density of the normal distribution as $n \to \infty$. Indeed, the probability density of the sum $\xi_1 + \xi_2$ is

$$p_2(x) = \int_{-\frac{1}{e}}^{\frac{1}{e}} p(z)\,p(x - z)\,dz.$$

We shall consider only those values of the argument $x$ which are near the point $x = 0$ (in particular we shall suppose that $|x| < 1/e$), and for the sake of definiteness we confine ourselves to considering only positive values of $x$. Under these conditions,

$$p_2(x) \geqslant \int_{-x}^{x} p(z)\,p(x - z)\,dz.$$

Since the minimum of the function $p(x - z)$ in the interval $0 \leq |z| \leq x$ is attained at the point $z = 0$,

$$p_2(x) \geqslant \frac{1}{2x|\log^2 x|} \int_{-x}^{x} \frac{1}{2|z|\log^2|z|} dz = \frac{1}{2x|\log^3 x|}.$$

In exactly the same way, it is easily seen that the probability density $p_3(x)$ of the sum $\xi_1 + \xi_2 + \xi_3$ satisfies the inequality

$$p_3(x) > \frac{c_3}{|x|\log^4|x|},$$

where $c_3 > 0$ is a constant,in a neighborhood of the point $x = 0$.

In general, the probability density $p_n(x)$ of the sum $\xi_1 + \xi_2 + \cdots + \xi_n$ satisfies the relation

$$p_n(x) > \frac{c_n}{|x \log^{n+1}|x||} \qquad (c_n > 0)$$

in a neighborhood of the point $x = 0$.

Thus for every $n$ the function $p_n(x)$ is infinite for $x = 0$. This means that $p_n(x)$ cannot converge to the density of the normal distribution under any normalization.

The example above compels us to search for sufficiently general conditions under which the probability densities of the sums converge to the density of the limit distribution.

THEOREM 1.* *Let the random variables of the sequence*

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

*be independent, identically distributed, and have the probability density* $p(x)$.† *If*

1) *for a certain* $m \geq 1$ *the probability density* $p_m(x)$ *of the sum* $\xi_1 + \xi_2 + \cdots + \xi_m$ *is integrable in the rth power* $(1 < r \leq 2)$ [*as we say, belongs to the space* $L^{(r)}$], *and*

2) $\int x^2 p(x)\, dx < +\infty,$

* *Translator's note.* Theorems 1, 2 and the Theorem in § 47 are simplified versions given in the Hungarian translation by I. Foldes (Budapest, 1951) of the Russian book.

† It is sufficient to assume that the density $p_m(x)$ of the sum $\xi_1 + \xi_2 + \cdots + \xi_m$ exists for some $m \geq 1$.

*then the relation*

$$\sigma \sqrt{n}\, p_n(\sigma \sqrt{n}\, x) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad (n \to \infty)$$

*holds uniformly with respect to $x$ in the interval* $(-\infty < x < \infty)$, *where*

$$\sigma^2 = \int x^2 p(x)\, dx.$$

Before turning to the proof, we shall make two simple remarks, which give us some idea about the class of distributions for which the conditions of the theorem are satisfied.

*Remark* 1. If for a certain $m \geq 1$ the density $p_m(x)$ satisfies a Lipschitz condition of order $\mu$ $(0 < \mu \leq 1)$, then the first condition of the theorem is satisfied. First of all, the fact that $p_m(x)$ satisfies a Lipschitz condition implies that the function $p_m(x)$ is bounded, and consequently that

$$\int p_m^r(x)\, dx < \infty.$$

for every $r > 0$.

*Remark* 2. If the probability density $p(x)$ is a function of bounded variation, then the first condition of the theorem is again satisfied. To prove this we shall show that under this assumption the function $p_2(x)$ will satisfy a Lipschitz condition. In fact, applying to the integral

$$p_2(x) = \int p(z)\, p(x-z)\, dz$$

the formula of integration by parts, we find that

$$p_2(x) = \int F(z)\, dp(x-z),$$

where

$$F(x) = \int_{-\infty}^{x} p(z)\, dz.$$

In exactly the same way,

$$p_2(x+h) = \int F(z+h)\, dp(x-z).$$

Thus

$$|p_2(x+h) - p_2(x)| \leqslant \int |F(z+h) - F(z)|\, |dp(x-z)|. \qquad (1)$$

Since the function $F(x)$ has a bounded derivative [otherwise the variation of $p(x)$ would be infinite] $F(x)$ satisfies a Lipschitz condition of order one. The inequality (1) shows that the density $p_2(x)$ also satisfies a Lipschitz condition of order one.* Q.E.D.

* We recall that the total variation of the function $p(x)$ is equal to $\int |dp(x)|$.

*Proof of Theorem* 1.

Let
$$f(t) = \int e^{itx} p(x)\,dx,$$
then
$$f^n(t) = \int e^{itx} p_n(x)\,dx.$$

It is known that the first condition of our theorem implies the integrability of the function $|f(t)|^n$ for all $n$ satisfying the inequality
$$n > \frac{mr}{r-1}\,^{*}$$
(see, for example, Titchmarsh [92], Theorem 96, p. 74). Therefore for all $n$ satisfying this inequality, we can write the inversion formula as follows:
$$2\pi p_n(x) = \int e^{-itx} f^n(t)\,dt.$$

We put $B_n = \sigma\sqrt{n}$. It is easily seen that
$$2\pi B_n p_n(B_n x) = \int e^{-izx} f^n\left(\frac{z}{B_n}\right) dz. \tag{2}$$
Since
$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \frac{1}{2\pi} \int e^{-izx-\frac{z^2}{2}} dz,$$
to prove the theorem it is sufficient to show that as $n \to \infty$
$$R_n = \int e^{-izx}\left[f^n\left(\frac{z}{B_n}\right) - e^{-\frac{z^2}{2}}\right] dz \to 0$$
uniformly with respect to $x$ $(-\infty < x < \infty)$. To this end we represent $R_n$ as the sum of four integrals:
$$I_1 = \int_{-A}^{A} e^{-izx}\left[f^n\left(\frac{z}{B_n}\right) - e^{-\frac{z^2}{2}}\right] dz, \quad I_2 = -\int_{|z|>A} e^{-izx-\frac{z^2}{2}} dz,$$
$$I_3 = \int_{A \leqslant |z| \leqslant \epsilon B_n} e^{-izx} f^n\left(\frac{z}{B_n}\right) dz, \quad I_4 = \int_{|z|>\epsilon B_n} e^{-izx} f^n\left(\frac{z}{B_n}\right) dz,$$
where the numbers $A > 0$ and $\epsilon > 0$ will be chosen later.

Since according to the second condition of the theorem $F(x)$ belongs to the domain of attraction of the normal law, whatever the constant $A$ may be
$$I_1 \to 0$$
as $n \to \infty$, uniformly with respect to $x$ $(-\infty < x < \infty)$.

* *Translator's note.* In the original, $m$ is missing from the formula.

By choosing $A$ sufficiently large, $|I_2|$ can be made as small as we wish.

By the second condition of the theorem the function $f(t)$ has continuous first and second derivatives, and

$$\left[\frac{d}{dt}f(t)\right]_{t=0}=0,\quad \left[\frac{d^2}{dt^2}f(t)\right]_{t=0}=-\sigma^2.$$

Hence, in a neighborhood of the point $t = 0$,

$$f(t)=1-\frac{\sigma^2t^2}{2}+o(t^2).$$

If $\epsilon > 0$ is sufficiently small, then for $|t| \leq \epsilon$ the remainder term can be made less than $\sigma^2t^2/4$ in modulus. Thus for $|t| \leq \epsilon$,

$$|f(t)|\leqslant 1-\frac{\sigma^2t^2}{4}\leqslant e^{-\frac{\sigma^2t^2}{4}}.$$

From this it follows that

$$|I_3|\leqslant \int\limits_{A\leqslant|z|\leqslant\epsilon B_n}\left|f^n\left(\frac{z}{B_n}\right)\right|dz<2\int\limits_A^\infty e^{-\frac{n\sigma^2t^2}{4B_n^2}}dt=2\int\limits_A^\infty e^{-\frac{t^2}{4}}dt.$$

Thus if $A$ is sufficiently large, $|I_3|$ can also be made as small as we wish. Since $|f(t)| \neq 1$ for $t \neq 0$ and $f(t) \to 0$ as $t \to \infty$ (as characteristic function of an absolutely continuous distribution function), it is possible to find a $c > 0$ such that $|f(t)| < e^{-c}$ for $|t| > \epsilon$. Let $\beta > mr/(r-1)$ be a constant, then

$$|I_4|\leqslant 2e^{-(nc-\beta)}\int\limits_{\epsilon B_n}^{\infty}\left|f\left(\frac{t}{B_n}\right)\right|^{\beta}dt.$$

Since the integral on the right side of the inequality converges, as $n \to \infty$

$$I_4\to 0.$$

Q.E.D.

If we make use of the lemma which will be proved in § 50, then it is possible to generalize the theorem above as follows:

THEOREM 2.* *Let the random variables*

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

*be mutually independent, identically distributed, and have the probability density* $p(x)$. *If*

1) *for a certain* $m \geq 1$ *the probability density* $p_m(x)$ *of the sum*

$$\xi_1+\xi_2+\cdots+\xi_m$$

*is integrable in the rth power* $(1 < r \leq 2)$,

---

* *Translator's note.* See the note to Theorem 1.

2) *the function* $F(x) = \int_{-\infty}^{x} p(z)\,dz$ *belongs to the domain of attraction of the stable law* $\Psi(x)$, *the characteristic function of which is defined by the formula* (1) of § 34 ($\alpha < 2$), *then the relation*

$$B_n p_n(B_n x + A_n) - p(x;\ \alpha,\ \beta,\ \gamma,\ c) \to 0 \quad (n \to \infty)$$

*holds uniformly with respect to* $x$ *in the interval* ($-\infty < x < \infty$), *where* $p(x; \alpha,\beta,\gamma,c) = \Psi'(x)$ *and the constants* $A_n$ *and* $B_n$ *are such that*

$$\mathsf{P}\left\{\frac{\xi_1 + \xi_2 + \ldots + \xi_n - A_n}{B_n} < x\right\} \to \Psi(x) \quad (n \to \infty).$$

## § 47. Improvement of the Limit Theorem for Densities

We now make somewhat stronger assumptions than in the preceding section, but on the other hand we obtain an expansion of the density $p_n(x)$ similar to that which was found for the distribution functions in § 45.

THEOREM.* *Let the random variables of the sequence*

$$\xi_1,\ \xi_2,\ \ldots,\ \xi_n,\ \ldots$$

*be independent, identically distributed, and have the probability density* $p(x)$. *If*

1) *for a certain* $m \geq 1$ *the probability density of the sum*

$$\xi_1 + \xi_2 + \ldots + \xi_m$$

*is integrable in the rth power* ($1 < r \leq 2$),

2) *for some* $k \geq 3$ ($k$ *an integer*)

$$\int |x|^k p(x)\,dx < \infty,$$

*then*

$$B_n p_n(xB_n) = \varphi(x) + \sum_{s=1}^{k-2} \frac{1}{n^{\frac{s}{2}}} P_s(-\varphi) + o\left(n^{-\frac{k-2}{2}}\right)$$

*uniformly with respect to* $x$ ($-\infty < x < \infty$), *where*

$$P_s(-\varphi) = \frac{d}{dx} P_s(-\Phi)$$

*and*

$$B_n^2 = n \int x^2 p(x)\,dx.$$

*Proof.* We know that [see (2) of § 46]

$$2\pi B_n p_n(xB_n) = \int e^{-izx} f^n\left(\frac{z}{B_n}\right) dz.$$

---

* *Translator's note.* See the note to Theorem 1, § 46.

Since [see (2) of § 45]

$$P_s(it)\,e^{-\frac{t^2}{2}} = \int e^{itx} P_s(-\varphi)\,dx,$$

by the inversion formula

$$2\pi P_s(-\varphi) = \int e^{-itx-\frac{t^2}{2}} P_s(it)\,dt$$

and consequently

$$2\pi\left[e^{-\frac{x^2}{2}} + \sum_{s=1}^{k-2} \frac{1}{n^{\frac{s}{2}}} P_s(-\varphi)\right] = \int e^{-itx} g(t)\,dt,$$

where for the sake of brevity we have put

$$g(t) = e^{-\frac{t^2}{2}}\left[1 + \sum_{s=1}^{k-2} \frac{1}{n^{\frac{s}{2}}} P_s(it)\right].$$

Thus

$$R_n = 2\pi\left[B_n p_n(xB_n) - \left\{e^{-\frac{x^2}{2}} + \sum_{s=1}^{k-2} \frac{1}{n^{\frac{s}{2}}} P_s(-\varphi)\right\}\right]$$

$$= \int e^{-izx}\left[f^n\left(\frac{z}{B_n}\right) - g(z)\right] dz.$$

Let $T_{nk} = \dfrac{n}{8k\rho_k^{3/k}}$ and represent $R_n$ as the sum of the following three integrals:

$$I_1 = \int_{-T_{nk}}^{T_{nk}} e^{-izx}\left[f^n\left(\frac{z}{B_n}\right) - g(z)\right] dz,$$

$$I_2 = \int_{|z|>T_{nk}} e^{-ixz} g(z)\,dz, \quad I_3 = \int_{|z|>T_{nk}} e^{-ixz} f^n\left(\frac{z}{B_n}\right) dz.$$

By Theorem 1(b) of § 41,

$$|I_1| \leqslant \frac{\delta(n)}{n^{\frac{k-2}{2}}} \int_{-T_{nk}}^{T_{nk}} [\,|t|^k + |t|^{3k-3}]\, e^{-\frac{t^2}{4}} dt = o\left(n^{-\frac{k-2}{2}}\right).$$

For $|z| > T_{nk}$,

$$\left|f\left(\frac{z}{B_n}\right)\right| < e^{-c},$$

where $c > 0$. Hence *

$$|I_3| < e^{-(nc-\beta)} \int_{|z|>T_{nk}} \left|f^n\left(\frac{z}{B_n}\right)\right|^{\beta} dz = o\left(n^{-\frac{k-2}{2}}\right).$$

* The number $\beta$ is defined as in § 46 in the estimation of the integral $I_4$.

Finally, it is obvious that

$$|I_2| \leqslant \int\limits_{|x| > T_{nk}} e^{-\frac{x^2}{2}} \left[1 + \sum_{s=1}^{k-2} \frac{1}{n^{\frac{s}{2}}} \left| P_s(it) \right| \right] dx = o\left(n^{-\frac{k-2}{2}}\right).$$

The estimates obtained prove the theorem.

The particular case of the theorem just proved, in which it is assumed that the density $p(x)$ is of bounded variation, was proved by H. Cramér [19].

## CHAPTER 9

## LOCAL LIMIT THEOREMS FOR LATTICE DISTRIBUTIONS

### § 48. Statement of the Problem

If each summand $\xi_m$ can take only values of the form

$$x_s = sh + a,$$

then the sum

$$\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n$$

takes only values of the form

$$z_{ns} = sh + na,$$

i.e., the distribution function $F_n(z)$ of the sum $\zeta_n$ is constant in each half-open interval

$$z_{ns} < z \leqslant z_{n,s+1}.$$

It is impossible to approximate such a distribution function with continuous (not to say analytic) functions to within one-half of its maximum jump:

$$\mathfrak{P}_n(s) = F_n(z_{ns} + 0) - F_n(z_{ns}).$$

However, the real interest lies, of course, only in the study of the values of the function $F_n(z)$ at the points $z_{ns}$ themselves, i.e., the sums

$$F_n(z_{ns}) = F_n(z_{n,s-1} + 0) = \sum_{r < s} \mathfrak{P}_n(r),$$

where

$$\mathfrak{P}_n(r) = \mathsf{P}(\zeta_n = z_{nr}).$$

No less interest lies in the study of the probabilities $\mathfrak{P}_n(s)$ themselves, the probabilities with which the sums $\zeta_n$ take different possible values $z_{ns}$. Moreover, it seems most natural to investigate the asymptotic behavior of the probabilities $\mathfrak{P}_n(s)$ as the primary problem. As is done, for example, in elementary textbooks of the theory of probability, the local theorem of Laplace is proved first and the integral theorem of Laplace is deduced from it as a consequence. It is possible to deduce in a similar way the integral theorem of § 43 from the local Theorem 2 of § 51, to be proved later. Generally speaking, such a method of proving integral theorems on the basis of local theorems may result in some loss in the precision of the estimates of the remainder terms as compared with direct methods of proof. However, for the first orientation in the problem the approach from the direction of local theorems seems preferable. This whole chapter

will be devoted to the asymptotic estimation of the probabilities $\mathfrak{P}_n(s)$ in the case of identically distributed independent lattice summands $\xi_m$.

In § 14 we have proved that every lattice distribution has a completely determined maximum span $h_0$. From an arbitrary span $h$ of the given distribution, the maximum span $h_0$ is obtained as follows. Form all possible differences

$$s' - s''$$

of the indices $s'$ and $s''$ for which (with the span $h$) the probabilities

$$\mathfrak{P}_1(s') = \mathsf{P}(\xi_m = s'h + a) \text{ and } \mathfrak{P}_1(s'') = \mathsf{P}(\xi_m = s''h + a)$$

are positive; find the greatest common divisor $\omega$ of all these differences and put

$$h_0 = \omega h.$$

In accordance with this, in order that the span $h$ itself be maximum, the following condition is necessary and sufficient.

($\omega$). The greatest common divisor of all the differences $s' - s''$ for which both probability $\mathfrak{P}_1(s')$ and probability $\mathfrak{P}_2(s'')$ are positive, is equal to one.

Naturally, it is sufficient to consider for each lattice distribution its maximum span, i.e., it is possible to confine our consideration to the case where the condition ($\omega$) is satisfied.

On the other hand, for any given span $h$ the transformation

$$\xi'_m = \frac{\xi_m - a}{h}, \quad \zeta'_n = \frac{\zeta_n - na}{h}$$

allows us to reduce the general case to the case in which the span is equal to *one*, while the constant $a$ is equal to *zero*. Combining the last two remarks, we see that in essence *it is sufficient* to consider the case of summands $\xi_m$, taking only integral values $s$ with

$$\mathfrak{P}_1(s) = \mathsf{P}(\xi_m = s),$$

subject to the condition ($\omega$).

## § 49. A Local Theorem for the Normal Limit Distribution

In accordance with § 48, we assume that the random variables $\xi_n$ take only integral values. The sum

$$\zeta_n = \xi_1 + \xi_2 + \dots + \xi_n$$

can also take only integral values. We put $\mathsf{P}\{\zeta_n = k\} = \mathfrak{P}_n(k)$. It is clear that for every $n$

$$\sum_{k=-\infty}^{\infty} \mathfrak{P}_n(k) = 1.$$

We assume also that

$$\mathsf{M}\xi_k = a, \ \mathsf{D}^2\xi_k = \sigma^2$$

and introduce the notation

$$z_{nk} = \frac{k - A_n}{B_n} = \frac{k - an}{\sigma\sqrt{n}}. \tag{1}$$

The object of this section is to prove the following proposition:

THEOREM. *Let the independent identically distributed random variables* $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ *take only integral values and have finite mathematical expectation* $a$ *and variance* $\sigma^2 \neq 0$. *In order that the relation*

$$B_n \mathfrak{P}_n(k) - \frac{1}{\sqrt{2\pi}} e^{-\frac{z_{nk}^2}{2}} \to 0 \qquad (n \to \infty)$$

*hold uniformly with respect to* $k$ *in the interval* $-\infty < k < \infty$, *it is necessary and sufficient that the greatest common divisor of the differences of all the values of* $\xi_n$ *taken with positive probabilities be equal to one* [Condition ($\omega$) of § 48].

The particular case of this theorem for variables $\xi_n$ taking only two values 0 and 1 with probabilities respectively equal to $p \neq 0$ and $q = 1 - p \neq 0$, forms the content of the classical local theorem of de Moivre-Laplace.

*Proof.* The characteristic function of the sum $\zeta_n$ is

$$f^n(t) = \mathsf{M}e^{it\zeta_n} = \sum_{k=-\infty}^{\infty} e^{itk}\mathfrak{P}_n(k).$$

Consequently, $\mathfrak{P}_n(k)$ can be calculated by the formula for Fourier coefficients:

$$2\pi\,\mathfrak{P}_n(k) = \int_{-\pi}^{\pi} f^n(t)\,e^{-itk}\,dt.$$

By (1)

$$k = z_{nk}B_n + A_n = zB_n + A_n$$

(in what follows we shall write $z$ instead of $z_{nk}$, omitting the indices); hence

$$2\pi\mathfrak{P}_n(k) = \int_{-\pi}^{\pi} e^{-itB_nz - itA_n} f^n(t)\,dt = \int_{-\pi}^{\pi} e^{-itB_nz} f^{*n}(t)\,dt,$$

where we have put

$$f^*(t) = e^{-ita}f(t).$$

Finally, making the substitution $x = tB_n$, we find that

$$2\pi B_n\mathfrak{P}_n(k) = \int_{-\pi B_n}^{\pi B_n} e^{-ixz} f^{*n}\left(\frac{x}{B_n}\right)dx.$$

Moreover, we know that

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = \frac{1}{2\pi} \int e^{-izx - \frac{x^2}{2}} dx.$$

Our problem consists in proving that as $n \to \infty$, the difference

$$R_n = 2\pi \left[ B_n \mathfrak{P}_n(k) - \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \right]$$

tends to 0 uniformly with respect to $k$ in the interval $(-\infty < k < \infty)$. To this end we represent $R_n$ as the sum of four integrals:

$$R_n = I_1 + I_2 + I_3 + I_4,$$

where

$$I_1 = \int_{-A}^{A} e^{-izx} \left[ f^{*n}\left(\frac{x}{B_n}\right) - e^{-\frac{x^2}{2}} \right] dx,$$

$$I_2 = \int_{A \leqslant |x| < \epsilon B_n} e^{-izx} f^{*n}\left(\frac{x}{B_n}\right) dx, \quad I_3 = \int_{\epsilon B_n < |x| < \pi B_n} e^{-izx} f^{*n}\left(\frac{x}{B_n}\right) dx,$$

$$I_4 = -\int_{|x| > A} e^{-izx - \frac{x^2}{2}} dx.$$

Here $A$ and $\epsilon$ are positive numbers which will be chosen later.

Since

$$|I_4| \leqslant 2 \int_{A}^{\infty} e^{-\frac{x^2}{2}} dx \leqslant \frac{2}{A} \int_{A}^{\infty} x e^{-\frac{x^2}{2}} dx = \frac{2}{A} e^{-\frac{A^2}{2}},$$

by choosing $A$ sufficiently large we can make the integral $I_4$ as small as we wish.

According to Corollary 2 to Theorem 5 of § 14 and the assumption that the maximum span is equal to one, for every given $\epsilon > 0$ it is possible to find a $c > 0$ such that for $\epsilon B_n \leq |x| \leq \pi B_n$

$$\left| f^{*n}\left(\frac{x}{B_n}\right) \right| \leqslant e^{-nc}.$$

Therefore,

$$|I_3| \leqslant e^{-nc} \int_{\epsilon B_n \leqslant |x| \leqslant \pi B_n} dx < 2\pi B_n e^{-nc} = 2\pi\sigma \sqrt{n}\, e^{-nc},$$

and consequently for fixed $\epsilon > 0$ the integral $I_3$ tends to zero as $n \to \infty$, uniformly with respect to $z$ $(-\infty < z < \infty)$.

Since the function $F(x)$, having a finite second moment, belongs to the domain of normal attraction of the normal law,

$$f^{*n}\left(\frac{x}{B_n}\right) \Rightarrow e^{-\frac{x^2}{2}}$$

and consequently for fixed $A$

$$I_1 \to 0$$

as $n \to \infty$, uniformly with respect to $z$.

To estimate the integral $I_2$ we remark that since $F(x)$ has a finite second moment we can write the following expansion for the function $f^*(t)$ in a sufficiently small neighborhood of the point $t = 0$:

$$\log f^*(t) = t\left[\frac{d}{dt}\log f^*(t)\right]_{t=0} + \frac{t^2}{2}\left[\frac{d^2}{dt^2}\log f^*(t)\right]_{t=0} + o(t^2)$$
$$= -\frac{t^2\sigma^2}{2} + o(t^2).$$

Thus, in the neighborhood of the point $t = 0$,

$$f^*(t) = e^{-\frac{t^2\sigma^2}{2} + o(t^2)}.$$

From this it follows that for sufficiently small $\epsilon > 0$ the inequality

$$\left|f^*(t)\right| \leqslant e^{-\frac{t^2\sigma^2}{4}}$$

holds in the interval $|t| \leq \epsilon$. Now

$$|I_2| \leqslant 2\int_A^{\epsilon B_n} \left|f^*\left(\frac{t}{B_n}\right)\right|^n dt \leqslant 2\int_A^{\epsilon B_n} e^{-\frac{t^2}{4}}\, dt < 2\int_A^{\infty} e^{-\frac{t^2}{4}} dt < \frac{4}{A} e^{-\frac{A^2}{4}}.$$

We see that by choosing $\epsilon$ sufficiently small, and $A$ sufficiently large, it is possible to make $I_2$ and $I_4$ as small as we wish (their estimates do not depend on $n$). The integrals $I_1$ and $I_3$ tend to zero as $n \to \infty$, whatever $A$ may be. Hence it follows that the whole sum $I_1 + I_2 + I_3 + I_4$ becomes as small as we wish for sufficiently large $n$, which proves the sufficiency of the conditions of our theorem.

The necessity of the conditions of the theorem is obvious from the fact that if the greatest common divisor $\omega$ of the differences of the possible values of $\xi_n$ is different from one, then the possible values of $\zeta_n$ $(n = 1,2, \ldots)$ will contain systematic gaps: the difference between two consecutive possible values of the sum $\zeta$ cannot be less than $\omega$.

## § 50. A Local Limit Theorem for Non-Normal Stable Limit Distributions

We assume that the distribution function $F_1(x)$ of the identically distributed independent summands $\xi_m$, taking only integral values, belongs to the domain of attraction of the stable law $G(y)$. Moreover, let $A_n$ and $B_n$ be constants for which

$$F_n(B_n y + A_n) = \mathsf{P}\left(\frac{\zeta_n - A_n}{B_n} < y\right) \Rightarrow G(y).$$

It is natural to raise the question of the applicability of the corresponding local limit formula

$$B_n\mathfrak{P}_n(k) = B_n \mathsf{P}(\zeta_n = k) \sim g\left(\frac{k - A_n}{B_n}\right),$$

where

$$g(y) = G'(y)$$

is the density corresponding to the distribution function $G(y)$.

The theorem of § 49 gives the answer to this question in the case where

$$g(y) = \varphi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}.$$

$$A_n = 0, \quad B_n = \sqrt{n\mathsf{D}^2\xi_m}.$$

Naturally the answer to the question in the case of an arbitrary normal law

$$g(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-a)^2}{2\sigma^2}}$$

with constants

$$A_n = n(\mathsf{M}\xi_m - a), \quad B_n = \frac{\sqrt{n\mathsf{D}^2\xi_m}}{\sigma}$$

can be reduced to the same theorem. The case of "non-normal" attraction to the normal law (see § 35) remains an unfinished study (B. V. Gnedenko [43]). The complete solution of the problem in the case of a non-normal law $G(y)$ is given by the following theorem (B. V. Gnedenko [45]):

THEOREM. *Let the independent identically distributed summands $\xi_m$ take only integral values, let $g(y)$ be the density of a certain non-normal stable distribution, and let $A_n$ and $B_n$ be certain constants. In order that the relation*

$$B_n\mathfrak{P}_n(k) - g\left(\frac{k - A_n}{B_n}\right) \to 0$$

*hold uniformly with respect to $k$, it is necessary and sufficient that the following two conditions be satisfied simultaneously:*

1) $F_n(B_ny + A_n) \Rightarrow G(y)$.
2) *Condition* ($\omega$) *of* § 48.

As examples we shall cite two particular cases of this theorem for specified distributions and specially chosen normalizing constants $B_n$.

1. (Local theorem for the Cauchy law.) *For a certain constant $\sigma > 0$ the relation*

$$\sigma n\mathfrak{P}_n(k) - \frac{1}{\pi\left(1 + \frac{k^2}{\sigma^2 n^2}\right)} \to 0 \qquad (n \to \infty)$$

*is satisfied uniformly with respect to $k$ $(-\infty < k < \infty)$ if and only if*

1) $\lim\limits_{x\to-\infty} -xF(x) = \lim\limits_{x\to\infty} x(1-F(x)) = \sigma$,*

2) *condition* $(\omega)$ *is satisfied.*

2. *For a certain constant $\sigma > 0$ the relation*

$$\sigma n^2 \mathscr{P}_n(k) - g\left(\frac{k}{\sigma n^2}\right) \to 0 \qquad (n\to\infty),$$

*where*

$$g(x) = \begin{cases} 0 & \text{for } x \leqslant 0, \\ \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2x}} x^{-\frac{3}{2}} & \text{for } x > 0 \end{cases}$$

*is satisfied uniformly with respect to $k$ $(-\infty < k < \infty)$ if and only if*

1) $\lim\limits_{x\to-\infty} \sqrt{|x|}\, F(x) = \lim\limits_{x\to\infty} \sqrt{x}\,(1-F(x)) = \sqrt{\sigma}$,*

2) *condition* $(\omega)$ *is satisfied.*

The proof of the theorem formulated above proceeds in the main on the arguments by which the Theorem of § 49 was proved. First of all, putting

$$f^*(t) = f(t)\, e^{-it\frac{A_n}{n}},$$

we obtain the equation ‡

$$B_n \cdot 2\pi \mathscr{P}_n(k) = \int_{-\pi B_n}^{\pi B_n} e^{-izt} f^{*n}\left(\frac{t}{B_n}\right) dt.$$

Furthermore, by the inversion formula,

$$g(z) = \frac{1}{2\pi} \int e^{-izt} v(t)\, dt,$$

where $\log v(t)$ is defined by formula (1) of § 34.

To estimate the difference,

$$R_n = 2\pi\,[B_n \mathscr{P}_n(k) - g(z)],$$

we represent $R_n$ as the sum of four integrals

$$R_n = I_1 + I_2 + I_3 + I_4,$$

---

* With respect to the condition 1) see Theorem 2 of § 35.

‡ As in § 49,

$$z = z_{nk} = \frac{k - A_n}{B_n}.$$

where

$$I_1 = \int_{-A}^{A} e^{-izt}\left[f^{*n}\left(\frac{t}{B_n}\right) - v(t)\right] dt,$$

$$I_2 = \int_{A \leqslant |t| \leqslant \epsilon B_n} e^{-izt} f^{*n}\left(\frac{t}{B_n}\right) dt, \quad I_3 = \int_{\epsilon B_n \leqslant |t| \leqslant \pi B_n} e^{-izt} f^{*n}\left(\frac{t}{B_n}\right) dt,$$

$$I_4 = -\int_{|t| > A} e^{-izt} v(t)\, dt.$$

As proved in the preceding theorem, the integrals $I_1$, $I_3$, and $I_4$ are as small as we wish for sufficiently large $n$ and $A$, no matter how small the previously chosen $\epsilon > 0$ may be. (For the estimation of $I_4$ we must take into account the fact that by (1) of § 34

$$\int |v(t)|\, dt < \infty.)$$

We shall now prove that by choosing $\epsilon$ sufficiently small and $A$ sufficiently large, it is possible also to make the integral $I_2$ as small as we wish. For this purpose we shall make use of the following lemma:

LEMMA. *If the distribution function $F(x)$ belongs to the domain of attraction of the stable law with the characteristic exponent $\alpha$ $(0 < \alpha < 2)$, then a constant $c_0 > 0$ can be found such that in a sufficiently small neighborhood of the point $t = 0$ the inequality*

$$|f(t)| < e^{-c_0 \tilde{\chi}\left(\frac{1}{|t|}\right)}, \tag{1}$$

*holds,* where*

$$\tilde{\chi}(x) = 1 - \tilde{F}(x) + \tilde{F}(-x)$$

*and*

$$\tilde{F}(x) = F(x) * [1 - F(-x + 0)].$$

*Proof.* By the assumption about the function $F(x)$ it is clear that $\tilde{F}(x)$ belongs to the domain of attraction of the stable law $\tilde{G}(x) = G(x) * [1 - G(-x)]$, for which the characteristic function is

$$\tilde{v}(t) = e^{-2c|t|^\alpha}.$$

* As an exercise, the reader may verify that the inequality

$$|f(t)| \leqslant e^{-c_0 \chi\left(\frac{1}{t}\right)},$$

also holds, where $\quad \chi(x) = 1 - F(x) + F(-x).$

Moreover, the normalizing constants $B_n$ for $F(x)$ and $\tilde{F}(x)$ may be chosen to be the same. Therefore, according to Theorem 2 of § 35, for every $u > 0$ and $x \to \infty$

$$\frac{\tilde{\chi}(ux)}{\tilde{\chi}(x)} \to \frac{1}{u^{\alpha}}.$$

Now

$$\tilde{f}(t) = \int \cos tx \, d\tilde{F}(x),$$

Hence

$$1 - \tilde{f}(t) = \int (1 - \cos tx) \, d\tilde{F}(x).$$

For every $t$ and $x$ the inequality $1 - \cos tx \geq 0$ holds; hence

$$1 - \tilde{f}(t) \geqslant \int\limits_{\frac{\pi}{2|t|} \leqslant |x| \leqslant \frac{3\pi}{2|t|}} (1 - \cos tx) \, d\tilde{F}(x).$$

But in the domain of integration,

$$1 \leqslant 1 - \cos tx \leqslant 2;$$

consequently,

$$1 - \tilde{f}(t) \geqslant \int\limits_{\frac{\pi}{2|t|} \leqslant |x| < \frac{3\pi}{2|t|}} d\tilde{F}(x) \geqslant^{*} \tilde{\chi}\left(\frac{\pi}{2|t|}\right) - \tilde{\chi}\left(\frac{3\pi}{2|t|}\right)$$

$$= \tilde{\chi}\left(\frac{1}{|t|}\right) \left[ \frac{\tilde{\chi}\left(\frac{\pi}{2|t|}\right)}{\tilde{\chi}\left(\frac{1}{|t|}\right)} - \frac{\tilde{\chi}\left(\frac{3\pi}{2|t|}\right)}{\tilde{\chi}\left(\frac{1}{|t|}\right)} \right].$$

Now let $|t| \leq \epsilon$ and let $\epsilon > 0$ be so small that

$$\frac{\tilde{\chi}\left(\frac{\pi}{2|t|}\right)}{\tilde{\chi}\left(\frac{1}{|t|}\right)} = \left(\frac{2}{\pi}\right)^{\alpha} + \omega_1, \quad \frac{\tilde{\chi}\left(\frac{3\pi}{2|t|}\right)}{\tilde{\chi}\left(\frac{1}{|t|}\right)} = \left(\frac{2}{3\pi}\right)^{\alpha} + \omega_2,$$

and for all $t$ ($|t| \leq \epsilon$)

$$\max(|\omega_1|, |\omega_2|) < \frac{1}{4}\left[\left(\frac{2}{\pi}\right)^{\alpha} - \left(\frac{2}{3\pi}\right)^{\alpha}\right] = c_0.$$

From this it follows that for $|t| \leq \epsilon$

$$1 - \tilde{f}(t) > 2c_0 \tilde{\chi}\left(\frac{1}{|t|}\right)$$

---

* *Translator's note.* In the original an equality sign is written, which is incorrect if $F(x)$ is discontinuous.

and so

$$\tilde{f}(t) < 1 - 2c_0\tilde{\chi}\left(\frac{1}{|t|}\right) \leqslant e^{-2c_0\tilde{\chi}\left(\frac{1}{|t|}\right)}.$$

Since

$$\tilde{f}(t) = |f(t)|^2,$$

the inequality obtained proves the lemma.

We can now turn to the estimation of the integral $I_2$. To this end we choose $\epsilon > 0$, so that the inequality (1) is satisfied in the interval $|t| \leq \epsilon$. Then for $|t| \leq \epsilon B_n$, we have

$$\left|f^*\left(\frac{t}{B_n}\right)\right|^n = \left|f\left(\frac{t}{B_n}\right)\right|^n \leqslant e^{-c_0 n\tilde{\chi}\left(\frac{B_n}{|t|}\right)}.$$

But for sufficiently large $n$ [see § 35, (9) and (10)],

$$n\tilde{\chi}\left(\frac{B_n}{|t|}\right) \sim c_1 |t|^\alpha. \qquad (c_1 > 0).^*$$

Consequently, in all cases, for sufficiently large $n$,

$$n\tilde{\chi}\left(\frac{B_n}{|t|}\right) \geqslant \frac{1}{2} c_1 |t|^\alpha.$$

Therefore for $|t| \leq \epsilon B_n$ and sufficiently large $n$,

$$\left|f\left(\frac{t}{B_n}\right)\right|^n \leqslant e^{-\frac{c_0 c_1}{2}|t|^\alpha}.$$

Now

$$|I_2| \leqslant \int_A^{\epsilon B_n} e^{-\frac{c_0 c_1}{2}|t|^\alpha}\, dt < \int_A^\infty e^{-\frac{c_0 c_1}{2}|t|^\alpha}\, dt.$$

The last integral can be made as small as we wish by the choice of $A$. This completes the proof of the theorem.

## § 51. Improvement of the Limit Theorem in the Case of Convergence to the Normal Distribution

We now turn to the improvement of the results of § 49. To do so we shall have to impose stronger conditions on the summands. Suppose the random variables $\xi_n$ can take only the values

$$x_s = a + sh \qquad (s = 0, \pm 1, \pm 2, \ldots),$$

---

* *Translator's note*. In the original the constant $c_1$ is missing. According to § 34, and using the notations there, $c_1$ is related to the $c$ occurring in $v(t)$ as follows:

$$c = c_1 L(\alpha) \cos\frac{\pi}{2}\alpha \text{ if } \alpha < 1;\ c = c_1\frac{\pi}{2} \text{ if } \alpha = 1;\ c = -c_1 M(\alpha)\cos\frac{\pi}{2}\alpha \text{ if } \alpha > 1.$$

where $h$ is the maximum span of the distribution.* The random variable

$$\eta_n = \frac{1}{B_n} \sum_{k=1}^{n} (\xi_k - \mathsf{M}\xi_k)$$

can take only the values

$$y = y_{ns} = \frac{h(s - n\bar{p})}{\sigma\sqrt{n}},$$

where

$$\bar{p} = \sum_{s=-\infty}^{\infty} sp_s \text{ and } p_s = \mathsf{P}\{\xi_k = a + sh\}.$$

From the fact that

$$f_n(t) = \sum_{s=-\infty}^{\infty} \mathfrak{P}_n(s) e^{ity_{ns}},$$

where $\mathfrak{P}_n(s) = \mathbf{P}\{\eta_n = y_{ns}\}$, we find that, as in the preceding section,

$$\mathfrak{P}_n(s) = \frac{1}{\tau\sigma\sqrt{n}} \int_{-\frac{1}{2}\tau\sigma\sqrt{n}}^{\frac{1}{2}\tau\sigma\sqrt{n}} f_n(t) e^{-ity_{ns}} dt.$$

As before, $\tau = 2\pi/h$. We shall now prove the following proposition (Esseen [26]):

THEOREM 1. *If the identically distributed lattice random variables $\xi_1$, $\xi_2, \ldots, \xi_n$ are independent and have finite absolute moments of the order $k$ $(k \geq 3)$ inclusive, then*

$$\mathfrak{P}_n(s) = \frac{h}{\sigma\sqrt{n}} \left\{\varphi(y_{ns}) + \sum_{\nu=1}^{k-2} \frac{1}{n^{\frac{\nu}{2}}} P_\nu(-\varphi(y_{ns}))\right\} + o\left(\frac{1}{n^{\frac{k-1}{2}}}\right).$$

*Here* $\varphi(y_{ns}) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{y_{ns}^2}{2}}$ *and $P_\nu(-\varphi)$ is defined as the function $P_\nu(-\Phi)$ in § 41 by substituting $\varphi$ for $\Phi$.*†

*Proof.* We have

$$\tau\sigma\sqrt{n}\,\mathfrak{P}_n(s) = I_1 + I_2, \tag{1}$$

---

* As already remarked in § 48, it is possible to confine ourselves to the case $h = 1$, as was done in §§ 49–50. We do not do this here for the sake of more convenient comparison with the results of §§ 43–44.

† *Translator's note.* $P_\nu(-\varphi)$ is obtained from the polynomial $P_\nu(-w)$ defined in § 38 by substituting $\varphi^{(r)}$ for $w^r$. Cf. (28) of § 38 and also the statement of the Theorem in § 47.

where

$$I_1 = \int_{-\frac{1}{2}\tau\sigma\sqrt{n}}^{\frac{1}{2}\tau\sigma\sqrt{n}} \left\{ f_n(t) - e^{-\frac{t^2}{2}} \left[ 1 + \sum_{\nu=1}^{k-2} \frac{1}{n^{\frac{\nu}{2}}} P_\nu(it) \right] \right\} e^{-iyt}\, dt$$

and

$$I_2 = \int_{-\frac{1}{2}\tau\sigma\sqrt{n}}^{\frac{1}{2}\tau\sigma\sqrt{n}} e^{-\frac{t^2}{2}} \left[ 1 + \sum_{\nu=1}^{k-2} \frac{P_\nu(it)}{n^{\frac{\nu}{2}}} \right] e^{-iyt}\, dt$$

(henceforth we shall write $y$ instead of $y_{ns}$).

We shall suppose that $T_{kn} < \tau\sigma\sqrt{n}/2$; otherwise the estimates will only be simplified. We can then write the equation

$$I_1 = I_3 + I_4, \tag{2}$$

where

$$I_3 = \int_{-T_{kn}}^{T_{kn}} \left\{ f_n(t) - e^{-\frac{t^2}{2}} \left[ 1 + \sum_{\nu=1}^{k-2} \frac{P_\nu(it)}{n^{\frac{\nu}{2}}} \right] \right\} e^{-iyt}\, dt$$

and

$$I_4 = \int\limits_{T_{nk} < |t| \leq \frac{1}{2}\tau\sigma\sqrt{n}} \left\{ f_n(t) - e^{-\frac{t^2}{2}} \left[ 1 + \sum_{\nu=1}^{k-2} \frac{P_\nu(it)}{n^{\frac{\nu}{2}}} \right] \right\} e^{-iyt}\, dt.$$

Using Theorem 1(b) of § 41, we find that

$$I_3 = o\left( \frac{1}{n^{\frac{k-2}{2}}} \right). \tag{3}$$

In $T_{nk} < |t| \leq \frac{1}{2}\tau\sigma\sqrt{n}$ there exists a constant $c > 0$ such that $|f_n(t)| < e^{-cn}$, hence it is not difficult to see that

$$I_4 = o\left( \frac{1}{n^{\frac{k-2}{2}}} \right). \tag{4}$$

But

$$\int\limits_{|t| > \frac{1}{2}\tau\sigma\sqrt{n}} e^{-\frac{t^2}{2}} \left[ 1 + \sum_{\nu=1}^{k-2} \frac{P_\nu(it)}{n^{\frac{\nu}{2}}} \right] e^{-ity}\, dt = o\left( \frac{1}{n^{\frac{k-2}{2}}} \right),$$

so that

$$I_2 = \int e^{-\frac{t^2}{2}} \left[ 1 + \sum_{\nu=1}^{k-2} \frac{P_\nu(it)}{n^{\frac{\nu}{2}}} \right] e^{-ity}\, dt + o\left( \frac{1}{n^{\frac{k-2}{2}}} \right). \tag{5}$$

Since

$$\int e^{-\frac{t^2}{2}} \left[ 1 + \sum_{\nu=1}^{k-2} \frac{P_\nu(it)}{n^{\frac{\nu}{2}}} \right] e^{-ity}\, dt$$

$$= \sqrt{2\pi} \left[ e^{-\frac{y^2}{2}} + \sum_{\nu=1}^{k-2} \frac{P_\nu(-\varphi(y))}{n^{\frac{\nu}{2}}} \right], \tag{6}$$

the equations (1)–(6) prove the theorem.

As a particular case of Theorem 1 for $k = 3$ we obtain:

THEOREM 2. *If the identically distributed lattice random variables* $\xi_1$, $\xi_2, \ldots, \xi_n$ *with maximum span* $h$ *and* $\mathbf{M}\xi_k = 0$ *are independent and have finite third moments, then*

$$\mathfrak{P}_n(s) = \frac{h}{\sigma\sqrt{n}} \varphi(y_{ns}) \left\{ 1 + \frac{1}{\sqrt{n}} \frac{\alpha_3}{\sigma^3} (y_{ns}^3 - 3y_{ns}) \right\} + o\left(\frac{1}{n}\right).$$

From this local theorem it is possible to obtain the integral theorem of § 43. As in the deduction of the integral theorem of de Moivre-Laplace from the local theorem, here in the infinite sum

$$F_n(z_{ns}) = \sum_{r<s} \mathfrak{P}_n(r)$$

the estimate of Theorem 2 is directly applied only to terms $\mathfrak{P}(r)$ with indices $r$ which are not far from $\bar{p}$, while the terms with large differences $|r - \bar{p}|$ are estimated specially.

From Theorem 1 for $k > 3$ it is possible to obtain corresponding estimates for $F_n(z)$ with a remainder term of the order

$$o\left( \frac{1}{n^{\frac{k-2}{2}}} \right),$$

We shall not give these estimates here (see Esseen [26]).

# APPENDIX I

## Notes on Chapter 1

*by* J. L. Doob

It is possible, although undesirable and unnatural, to write a book on the limit distributions of sums of independent random variables with essentially the content of this book, but without the use of the name *random variable* anywhere in the text. This can be done as follows. Let $x_1, \ldots, x_n$ be mutually independent random variables, with respective distribution functions $F_1, \ldots, F_n$, and respective characteristic functions $f_1, \ldots, f_n$. Then the distribution function of $s_n = x_1 + \cdots + x_n$ is the convolution of the $F_j$'s, and the characteristic function of $s_n$ is the product of the $f_j$'s. The standard procedure is to investigate the distribution of $s_n$ by means of its characteristic function. One can carry through this work with no reference to the $x_j$'s or $s_n$, by simply discussing the iterated convolutions of distribution functions, and the corresponding products of characteristic functions. In this book, since much of it is concerned only with characteristic functions, it would even be possible to phrase much of the material entirely in terms of characteristic functions, omitting reference both to random variables and distribution functions. Although this type of treatment is not uncommon in distribution theory, it would be undesirable in a large work of the present kind, since it would have been misleading to give the great quantity of material in this book without reference to the basic theory of probability which provides the context that gives the material its importance.

Although not much of the basic theory of probability is needed for this book, the authors quite properly judged a short outline to be necessary, since there is no reference book available in any language which covers it properly. It is an interesting fact that with all the research going on in the theory of probability, there is still no text (such as there are in great numbers, for example, in the theory of functions of a complex variable) which starts from the beginning, makes all the necessary definitions, with a proper discussion of each, and proves the basic theorems, thus assuring the student that only the analytical details peculiar to particular developments remain to hinder him. The point is not that there is no good book which covers this material, but that no book has even been written with this purpose in mind. This appendix is written to elaborate in more detail some aspects of the basic theory which appear in rather compact form in Chapter 1.

There is now essential unanimity among mathematicians working in probability that, for mathematical purposes, an event is a measurable point set, the probability of an event is the measure of the point set, a

random variable is a measurable function, and the expectation of a random variable is the integral of the function. It is not generally realized, however, that these basic definitions do not suffice to set up the theory of probability, but that considerable elaboration is necessary, along the following lines.

P1 *The various models for families of random variables and associated measure spaces must be treated.*

P2 *Conditional probabilities and expectations must be defined, and their properties established.*

P3 *The basic probability measure must be discussed, in terms of the desirability of suitable restrictions and canonical modifications.*

These three topics have not been listed in the order in which they would be treated in a systematic text, but in the order in which it will be convenient to treat them here. In the following we shall suppose, until we come to the discussion of P3, that the given probability measure satisfies only the restrictions ($\mu$1) and ($\mu$2) given in § 2. The convenience and necessity of further restrictions will be discussed later.

*Remarks on* P1. Random variables were defined above. Their existence, satisfying specified conditions, is a separate question, usually solved by the use of mathematical models.

For example, consider Theorem 4, § 21, whose statement begins: *Let*

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

*be a sequence of mutually independent random variables and let the distribution function of* $\xi_k$ *be* $F_k(x)$. The theorem is quite correct without an answer to the following question, but its importance is considerably enhanced by the fact that the answer is affirmative. The question is: "If the $F_k$'s are specified, is there a corresponding sequence of mutually independent random variables, defined on some measure space?" It would be unfortunate if the answer to this question depended on the particular distributions involved. Actually, there is always such a sequence of random variables, and the random variables can be taken as the coordinate variables of an infinite dimensional coordinate space. In fact, Kolmogorov showed [65] that a family of random variables $\{\xi_\alpha, \alpha \in T\}$ indexed in an arbitrary set $T$ can be defined if, for each finite index set $\alpha_1, \ldots, \alpha_n$, the $n$-variate distribution of $\xi_{\alpha_1}, \ldots, \xi_{\alpha_m}$ is specified, as long as these specified finite dimensional distributions are compatible. Compatibility means that, if $m < n$, the marginal distribution of $\xi_{\alpha_1}, \ldots, \xi_{\alpha_m}$ obtained from the above is exactly the specified distribution of $\xi_{\alpha_1}, \ldots, \xi_{\alpha_m}$. Kolmogorov defined the $\xi_\alpha$'s as random variables on a coordinate space with $\xi_\alpha$ the $\alpha$th coordinate variable, defining a measure of sets in this coordinate space in such a way that the $\xi_\alpha$'s have the desired distributions. In particular, if any family of random variables $\{\xi'_\alpha, \alpha \in T\}$ is *given*, the distributions of finite sets

of these random variables determine, according to this method, a measure in a coordinate space whose dimensionality is the cardinal number of $T$, and a family $\{\xi_\alpha, \alpha \in T\}$ of coordinate random variables of this space, with the property that, for any finite index set $\alpha_1, \ldots, \alpha_n$, the two sets of random variables

$$\xi'_{\alpha_1}, \ldots, \xi'_{\alpha_n}, \qquad \xi_{\alpha_1}, \ldots, \xi_{\alpha_n},$$

defined on different spaces, have the same $n$-variate distribution. Thus distribution problems involving the original variables can be stated in terms of the coordinate variables, and are frequently thereby simplified. For example, if there is a finite number $n$ of random variables in the given family, that is, if the index set $T$ contains only $n$ points, distribution problems are reduced to problems involving ordinary distributions in $n$-dimensional space. In the latter case, of course, the $n$-dimensional distribution is derived from the given random variables by the simple map described in § 2. The given $n$ random variables map the basic space of elementary events into a subset of $n$-dimensional space; this map is used to define a measure in $n$-dimensional space in such a way that the transformation between the two spaces is measure preserving. In the most general case, a family of random variables $\{\xi'_\alpha, \alpha \in T\}$ is mapped into a family $\{\xi_\alpha, \alpha \in T\}$. More precisely, the elementary event space on which the $\xi'_\alpha$'s are defined is mapped into the coordinate space on which the $\xi_\alpha$'s are defined. The function $\xi'_\alpha$ goes into the function $\xi_\alpha$, and the map preserves (probability) measure. It is desirable to have the closest possible relation between $\xi'_\alpha$'s and $\xi_\alpha$'s, and the hypothesis to be discussed below, that the basic probability measure is perfect, is a step in this direction.

Care must be taken to differentiate between the existence of a random variable defined on the given measure space and the existence of a distribution with specified properties. As an example, consider the definition of an infinitely divisible random variable, given in § 17. The random variable $\xi$ is there said to be infinitely divisible if, for every positive integer $n$, it can be expressed as the sum of $n$ independent identically distributed random variables. Note that this definition imposes two requirements, for each value of $n$: (i) the characteristic function of $\xi$ is to be the $n$th power of a characteristic function; (ii) the structure of the given measure space is complex enough to support $n$ identically distributed independent random variables with sum $\xi$. Actually the first (weaker) requirement is all that is usually of interest, and all that is treated in this book. It is true, although not proved in this book, that to every distribution whose characteristic function has the property (i) there is a probability measure space and a random variable defined on it whose distribution has this characteristic function and satisfies (ii). It is not true that if a random variable satisfies (i), it also satisfies (ii). In fact, if $\xi$ is a random variable with a Poisson distribution, (i) is satisfied, that is, $\xi$ has an infinitely divisible distribution,

but the representation $\xi = \xi_1 + \xi_2$, where $\xi_1$, $\xi_2$ are independent and have a common distribution, may or may not be possible, depending on the basic space. In fact, if $\xi$ has expectation 1, $\xi_1$ and $\xi_2$ must each have a Poisson distribution with expectation $\frac{1}{2}$, and if the basic space is a sequence of points, each point corresponding to a single value of $\xi$, it is easy to see that the stated representation is impossible. (The only probabilities that exist are sums of terms of the series $e^{-1} \sum_0^\infty 1/n!$, and the probability $e^{-\frac{1}{2}}$ that $\xi_1 = 0$ cannot be expressed in this form, as an elementary examination of possibilities shows.)

*Remarks on* P2. There is no reason to give a detailed discussion of conditional probabilities and expectations here. However, the following is an example of the difficulties that can arise, due to the fact that conditional probabilities and expectations are not uniquely defined. Suppose that conditional probabilities of sets have been defined, relative to some specified conditions which we omit. Then it is important to know when these conditional probabilities can be treated as ordinary probabilities, that is, as defining probability measures in terms of which integration yields conditional expectations. This is not always possible, but it is possible, and even trivially simple, if the condition is a specified one of positive probability. This special case is all that is needed for the present book, and is treated in Chapter 1.

*Remarks on* P3. We accept without argument the hypothesis that the measure in question is completely additive. This is, of course, not necessary as a requirement either philosophically or mathematically, but there is no present indication that there is any advantage in treating finitely additive measures.

The authors incorporate *completeness* of a measure [condition ($\mu$3) of § 2] as part of their measure definition. This condition is in no way relevant to the needs of the book, but it is frequently convenient in measure studies, and it is harmless in the following sense. A measure which is not already complete can be made complete by adding to the class of measurable sets every set which is not measurable but which has the property that there are two measurable sets of the same measure, one containing the set and the other contained in it. The nonmeasurable set is assigned as measure the common measure of these two measurable sets. This operation of completion not only leaves unchanged the measures of the given measurable sets, but adds to the class of measurable sets those and only those sets whose measures are uniquely determined by the given measure function if they are to be assigned measures.

The authors add a further condition to their measures; they are to be *perfect*. This condition is also quite unnecessary for the purposes of this book. The following remarks will, however, give some indication of the advantages to be gained from this restriction, which is certainly a con-

venience, although not a necessity. We shall first find a condition necessary and sufficient that a measure be perfect. This condition is really only a slight rephrasing of the definition, but gives insight into its significance.

Let $F$ be any distribution function of one variable. Define the outer measure of a linear set $A$ as inf $\sum_j [F(b_j) - F(a_j)]$, where $\bigcup_j [a_j,b_j)$ is any union of semiclosed intervals (including left-hand end points only) which covers $A$. This outer measure determines a measure in accordance with the usual Carathéodory method. We shall call the measurable linear sets obtained in this way *F-measurable*. This measure, sometimes called Lebesgue-Stieltjes measure based on $F$, is complete, and in fact the $F$ measure considered defined only on the linear Borel sets yields $F$ measure when completed as described above.

Now let $\xi$ be any random variable defined on a measure space satisfying conditions ($\mu$1), ($\mu$2), ($\mu$3) of § 2, and let $F$ be the distribution function of $\xi$. In the notation of § 2, the map $\xi' = \xi(u)$ maps the space $U$ of elementary events on which the given probability measure $\mu$ is defined, onto a subset of the line $U'$. A measure $\mu'$ of linear sets is defined by setting

$$\mu'(A) = \mu[\xi^{-1}(A)]$$

for every $A$ with the property that $\xi^{-1}(A)$ is $\mu$-measurable. This defines a measure on a certain Borel field $\mathfrak{M}_{\mu'}$ of $U'$ sets. If we make $A$ an interval here, we find that $\mu'$ and $F$ measures are the same on intervals, and therefore on all Borel sets. It is clear that $\mu'$ measure is complete. Hence, by definition of $F$ measure, the two measures are equal on the class of $F$-measurable sets. The two measures need not be identical, however, because $\mathfrak{M}_{\mu'}$ may contain sets which are not $F$-measurable.

THEOREM. *The $\mu$ measure is perfect if and only if, for every $\xi$, $\mu'$ and $F$ measures are identical.*

Suppose first that $\mu$ measure is perfect. This means that, if $A$ is $\mu'$-measurable, there is, for each positive integer $n$, an open set containing $A$ and of $\mu'$ measure at most $\mu'(A) + 1/n$. The intersection of a sequence of open sets obtained in this way for each value of $n$ is a Borel set $A_2$ containing $A$, with $\mu'(A) = \mu'(A_2)$. Applying this argument to the complement of $A$, we find a Borel set $A_1$, contained in $A$, the union of a sequence of closed sets, with $\mu'(A) = \mu'(A_1)$. Now $A_1$ and $A_2$ are Borel sets, and as such are $F$-measurable, with $F$ measures the same as their $\mu'$ measures. Since $F$ measure is complete, it follows that $A$ is also $F$-measurable, because it lies between two $F$-measurable sets of the same measure. Thus, if $\mu$ measure is perfect, $\mu'$ and $F$ measures are identical, in domains of definition and values. Conversely, if these two measures are identical for every choice of $\xi$, the $\mu'$ measure shares with $F$ measure the property that the $\mu'$ measure of a set is the lower limit of the $\mu' = F$ measures of contain-

ing open sets, and this property is the defining property of a perfect measure. (The fact that $F$ measure has the stated property will be found in any text which discusses Lebesgue-Stieltjes measures.)

Thus the essential significance of the hypothesis that the basic measure is perfect is that under the natural map of the space of elementary events into a line, discussed in § 2, determined by a random variable, or more generally under the map into Euclidean $n$-space determined by $n$ random variables, the only Euclidean space sets whose inverse images are measurable are the Borel sets and the sets obtained from these by completing the measure of Borel sets in $n$-space determined by the multivariate distribution function of the given $n$ random variables. Since this map is a useful tool in reducing problems involving $n$ random variables to those involving the coordinate variables in Euclidean $n$-space, it is sometimes convenient to have the simple relation just described.

Two examples will now be given illustrating on the one hand the convenience of the hypothesis that the basic probability measure is perfect, and on the other the fact that this hypothesis is by no means a necessity.

Let $\xi_1$, $\xi_2$ be random variables, defined on the same measure space. Then for the purposes of probability theory it is most useful to define $\xi_1$ to be independent of $\xi_2$ if, for every pair of linear Borel sets $A_1$, $A_2$,

$$\mathbf{P}\{\xi_1 \in A_1, \xi_2 \in A_2\} = \mathbf{P}\{\xi_1 \in A_1\}\, \mathbf{P}\{\xi_2 \in A_2\}. \tag{I-1}$$

(It is easy to see that this equation holds as stated if it holds whenever $A_1$ and $A_2$ are intervals, or even intervals with $-\infty$ as left-hand end point, and the condition is commonly stated with $A_1$ and $A_2$ of this form.) On the other hand, it would be natural from some points of view, and somewhat more elegant, to prescribe that (I-1) be true whenever $A_1$ and $A_2$ are sets of real numbers such that the right side of (I-1) is defined, that is, such that the two sets of elementary events involved on the right are measurable. This second definition, which is the one actually given by Gnedenko and Kolmogorov, is (almost trivially) equivalent to the first if the basic probability measure is perfect. The two definitions are not necessarily equivalent otherwise,* so that, without this restriction on the basic probability measure, the definition of independence given in this book would be modified to that given at the beginning of this paragraph.

As a second example, consider the problem of defining conditional probabilities and expectations. The details will not be given here, but if the basic measure is perfect, the problem mentioned under P2 has an affirmative answer. Without this hypothesis, conditional probabilities must be treated with somewhat greater care, but conditional probabilities nevertheless act about like probabilities for most purposes. For example,

---

* See B. Jessen, *Coll. Math.* **1**, 214–215 (1948), and J. L. Doob, *ibid.* pp. 216–217 for examples in which the two independence definitions are not equivalent.

the usual integral (expectation) inequalities associated with the names Bunyakovski, Cauchy, Euler, Hölder, Jensen, Minkowski, Schwarz, and so on are valid for conditional expectations just as they are for expectations. The point is that any problem involving a family of random variables can be reduced, even if the basic measure is not perfect, to a corresponding problem involving the coordinate variables of a coordinate space (see the discussion of P1) for which the problems associated with nonperfect measures do not arise. As is true in certain other problems, the important *fields* must be discovered and analyzed.

We conclude with a problem of a different type. In any discussion of continuous parameter families of random variables, as in the discussion of Brownian motion, it is desirable to make statements on the distributions of the superior limit of a nondenumerable collection of random variables, and on the continuity in the indexing parameter of sample functions of the family. The probabilities needed in such an analysis are not necessarily defined, that is, the corresponding classes of elementary events are not necessarily measurable, even if the basic probability measure is complete and perfect. Thus some standard method must be accepted, either for reinterpreting probabilities, or for modifying the basic measure space or measure defined on it, to make it possible to define the desired probabilities.*

---

* See J. L. Doob, *Stochastic Processes*, New York (1953).

## APPENDIX II

### NOTES ON § 32

Theorem 1 of § 32 needs amplification.

First, the so-called "distribution function" $V(x)$ there is not necessarily continuous to the left. In fact, it is not difficult to see that *if $F(x)$ is unimodal with vertex at $x = 0$, then the left derivative $F'_-(x)$ is continuous to the left, and the right derivative $F'_+(x)$ is continuous to the right for every $x \neq 0$* Hence at a point $x$ where $F'_-(x) \neq F'_+(x)$, the function $F(x) - xF'(x) = V(x)$ will be continuous to the left or to the right according as $F'(x)$ is taken to be the left or the right derivative at this point. This indetermination of $V(x)$ at discontinuity points is inessential, but it conflicts with the definition adopted in the book (see Chapter 1, § 6).

In the sufficiency part of Theorem 1, namely, in the statement "if $V(x) = F(x) - xF'(x)$ is a distribution function then $F(x)$ is unimodal," it is not clear what preliminary assumption is made on the distribution function $F(x)$. This is caused again by the ambiguity of $F'(x)$. Suppose that the preliminary assumption were that either $F'_-(x)$ exists everywhere or else $F'_+(x)$ exists everywhere; then the statement is clearly incorrect. To see this, we need only consider any (left-continuous) distribution which is a step function with more than one jump; then $F'_-(x) = 0$ everywhere and $F(x) - xF'_-(x)$ reduces to $F(x)$ itself, but $F(x)$ is not unimodal. In this example, of course, $F'_+(x)$ is infinite and $F(x)$ discontinuous at some point $x \neq 0$. The situation may be remedied by requiring that both $F(x) - xF'_-(x)$ and $F(x) - xF'_+(x)$ lie between $V(x - 0)$ and $V(x + 0)$ (both inclusive), for then $F'_-(x)$ and $F'_+(x)$ will be finite and $F(x)$ continuous for every $x \neq 0$. A simpler formulation is given in Theorem 1(b) below.

After these remarks we now give a precise version of Theorem 1. The term "distribution function" is used in the strict sense.

THEOREM 1. (a) *If the distribution function $F(x)$ is unimodal with vertex at $x = 0$, then there exists a distribution function $V(x)$ such that*

$$F(x) - xF'_-(x) = V(x)$$

*and*

$$F(x + 0) - xF'_+(x) = V(x + 0)$$

*for every $x$.* (A product $0 \cdot \infty$ is taken to be 0.)

(b) *Let the distribution function $F(x)$ be continuous except possibly at $x = 0$. Suppose that there is a denumerable set $D$ of points $x$ and a distribution function $V(x)$ such that if $x$ is not in $D$, the right or left derivative*

*$F'(x)$ (possibly different ones at different points) exists and satisfies the equation*

$$F(x) - xF'(x) = V(x);$$

*then $F(x)$ is unimodal.*

The proof of the theorem proceeds as in the text. In part (a) we use the italicized proposition given in the beginning of this Appendix. In part (b) we use the following theorem.*

*If $F$ is a continuous function on an interval $I$, and if at each point of this interval, except those of a denumerable set, one at least of the four Dini derivatives is equal to zero, then the function $F$ is constant on $I$.*

Theorem 3 of § 32, attributed to A. I. Lapin, asserts that "The composition of two unimodal distribution functions with vertex $O$ is unimodal with vertex $O$." The proof given of this theorem proceeds as follows. Let $F_i(x)$, $i = 1,2$ be unimodal with vertex $O$, and let $F_i'(x)$ denote the left derivative of $F_i(x)$. By Theorem 1, the two functions

$$V_i(x) = F_i(x) - xF_i'(x) \qquad (i = 1,2)$$

are distribution functions. If

$$F = F_1 * F_2 = F_2 * F_1,$$

then

$$F'(x) = (F_1 * F_2)'(x) = \frac{d}{dx}\int F_1(x - z)\, dF_2(z) = (F_1' * F_2)(x)$$
$$= (F_2' * F_1)(x) = \tfrac{1}{2}[(F_1' * F_2)(x) + (F_2' * F_1)(x)].$$

It follows that

$$F(x) - xF'(x) = \tfrac{1}{2}[F_1 * F_2 + F_2 * F_1] - x[F_1' * F_2 + F_2' * F_1]$$
$$= \tfrac{1}{2}[(F_1 - xF_1') * F_2 + (F_2 - xF_2') * F_1] = \tfrac{1}{2}[V_1 * F_2 + V_2 * F_1].$$

The right side of the last equation is a distribution function; hence by Theorem 1 $F(x)$ is unimodal with vertex $O$.

There are two errors in this proof. First, the equation $F' = F_1' * F_2$ holds only if $F_1(x)$ is continuous at $x = 0$, hence absolutely continuous on

---

* See Saks, *Theory of the Integral*, 2nd ed., Stechert, New York (1937), p. 272. If we assume, as is sufficient for our purpose, that at least one of $F'_-(x)$ and $F'_+(x)$ is equal to zero everywhere in $I$ except in a denumerable set of points, then the conclusion of the theorem can be proved by the following simple argument due to P. Erdös. If $F(x)$ is not constant we may suppose that there are two points $a$ and $b$ in $I$ such that $a < b$ and $F(a) < F(b)$. Let $F(a) < c < d < F(b)$ and consider the straight line $L(d)$ passing through $(a,c)$ and $(b,d)$. Let $x_0(d)$ be the supremum of the points in $(a,b)$ at which the curve $y = F(x)$ is strictly below $L(d)$. Obviously, neither $F'_-(x_0)$ nor $F'_+(x_0)$ can vanish. Two distinct values of $d$ correspond to two distinct values of $x_0(d)$. Hence there is a set of points in $(a,b)$ of the power of the continuum at which neither derivative vanishes.

$(-\infty, +\infty)$. Even in this case the differentiation under the integral needs justification. This can be done by Fubini's theorem, Lebesgue's convergence theorem, and the fact that $F_1' * F_2$ is continuous except possibly in a denumerable set. On the other hand, if Lapin's theorem were true for continuous $F_1$ and $F_2$, then it would be true also in general. This is easily seen if we write each $F_i$ as the sum of its jump at $x = 0$ and a continuous part to which the result applies.

The second error lies in the fact that

$$F_1 * F_2 - x(F_1' * F_2) = (F_1 - xF_1') * F_2$$

is not an identity. This error cannot be repaired to save the theorem.* In fact, the statement of the theorem itself is false, as the following trivial example shows. Let

$$F(x) = \begin{cases} x + \frac{1}{3} \text{ if } -\frac{1}{3} \le x \le \frac{2}{3}, \\ 0 \text{ otherwise.} \end{cases}$$

Then, according to the definition, $F$ is unimodal with vertex $O$, but $F * F$ is unimodal with (the unique) vertex $\frac{1}{3}$.

The statement of Theorem 3 remains false even if no specification is made about the vertices. We shall give an example in which $F$ is unimodal with vertex $O$ and absolutely continuous, but $F * F$ is not unimodal at all; in fact, its derivative (density function) is continuous and has two relative maxima.

Example.

$$p(x) = \begin{cases} 0 \text{ if } x < -\frac{1}{30}, \\ 5 \text{ if } -\frac{1}{30} \le x \le 0, \\ 1 \text{ if } 0 < x \le \frac{5}{6}, \\ 0 \text{ if } \frac{5}{6} < x. \end{cases}$$

$$F(x) = \int_{-\infty}^{x} p(z)\, dz.$$

The derivative of $F * F$ is then given by

$$p_2\,(x) = \int_{-\infty}^{\infty} p(x - z)p(z)\, dz.$$

* For a result which can be obtained by correcting the error, see K. L. Chung, Sur les distributions unimodales, *C. R. Acad. Sci. Paris*, **236**, 583-584 (1953).

Elementary computation gives the following explicit formula:

$$p_2(x) = \begin{cases} 0 & \text{if } x \leq -\frac{1}{15}, \\ 25x + \frac{5}{3} & \text{if } -\frac{1}{15} \leq x \leq -\frac{1}{30}, \\ -15x + \frac{1}{3} & \text{if } -\frac{1}{30} \leq x \leq 0, \\ x + \frac{1}{3} & \text{if } 0 \leq x \leq \frac{4}{5}, \\ -9x + \frac{25}{3} & \text{if } \frac{4}{5} \leq x \leq \frac{5}{6}, \\ \frac{5}{3} - x & \text{if } \frac{5}{6} \leq x \leq \frac{5}{3}, \\ 0 & \text{if } \frac{5}{3} \leq x. \end{cases}$$

Thus $p_2(x)$ has two relative maxima at $-\frac{1}{30}$ and $\frac{4}{5}$ with the values $\frac{5}{6}$ and $\frac{17}{15}$ respectively, and a minimum at 0 with the value $\frac{1}{3}$.

We may, if we wish, modify this example in such a way that $p(x)$ is continuous everywhere and attains its maximum at a unique point, while $p_2(x)$ still has more than one relative maximum. This follows from considerations of continuity.

Theorem 5 of § 32, attributed to Gnedenko, states that "all distribution functions belonging to the class $L$ are unimodal." The proof given depends essentially on the false Theorem 3 and therefore is not valid. Thus this interesting statement remains a conjecture. It is not even known whether all stable laws are unimodal. The only results known in this direction seem to be those of Wintner * which state that "the composition of two symmetrical unimodal distribution functions is symmetrical unimodal," and consequently that "all symmetrical stable laws are unimodal."

In this translation the original Theorems 3 and 5 are omitted and Theorem 4 is renumbered Theorem 3. As a consequence, proposition 1 of § 36 which states that "all stable laws are unimodal" is also omitted, and the subsequent propositions renumbered accordingly. Owing to this revision the material in § 32 will have no bearing on the rest of the book. However, we deem Theorems 1 and 2 of § 32 of sufficient independent interest to be included in the translation.

---

* A. Wintner, *Asymptotic Distributions and Infinite Convolutions.* Edwards Brothers, Ann Arbor, Michigan (1938), pp. 30 and 32.

# BIBLIOGRAPHY †

BAWLY, G. M.

[1] Ueber einige Verallgemeinerungen der Grenzwertsätze der Wahrscheinlichkeitsrechnung, Rec. Math. [Mat. Sbornik] N.S. **1** (**43**), 917–930 (1936).

[2] Über den lokalen Grenzwertsatz der Wahrscheinlichkeitsrechnung. Rev. Fac. Sci. Univ. Istanbul (N.S.) **2**, no. 2, 79–92 (1937).

BERNOULLI, JACQUES

[3] Ars Conjectandi. Basle, 1713.

BERNSTEIN, S. N.

*[4] On the law of large numbers. Soobshch. Kharkov. Matem. Obshchestva (2) **16**, 82–87 (1918).

[5] Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. Math. Ann. **97**, 1–59 (1926).

[6] Quelques remarques sur le théorème limite Liapounoff. C. R. (Doklady) Acad. Sci. URSS (N.S.), **24**, 3–8 (1939).

*[7] On P. L. Chebyshev's researches in the theory of probability. The Scientific Legacy of P. L. Chebyshev, vol. 1, 43–68 Izdat. Akad. Nauk SSSR, Moscow-Leningrad, 1945.

*[8] Theory of probability, 4th ed., GTTI, Moscow-Leningrad, 1946.

*[9] On a limit theorem in the theory of probability. Izvestiya Nauchno-Issled. Inst. Mat. Meh. Tomsk. Gos. Univ. **3**, no. 1, 174–190 (1946).

BERRY, A. C.

[10] The accuracy of the Gaussian approximation to the sum of independent variates. Trans. Amer. Math. Soc. **49**, 122–136 (1941).

BOBROV, A. A.

*[11] On the relative stability of sums in a particular case. Rec. Math. [Mat. Sbornik] N.S. **4** (**46**), 99–104 (1938).

*[12] On the relative stability of sums of positive random variables. Uchenye Zapiski Moskov. Gos. Univ. Matematika **15**, 191–202 (1939).

BRUNS, H.

[13] Wahrscheinlichkeitsrechnung und Kollektivmasslehre. Teubner, Leipzig-Berlin, 1906.

CAUCHY, A. L.

[14] Sur les résultats moyens d'observations de même nature et sur les résultats les plus probables. C. R. Acad. Sci. Paris **37**, 198–206 (1853).

CHARLIER, C. V. L.

[15] Ueber das Fehlergesetz. Ark. Mat. Astr. Fiz. **2**, no. 8 (1905–1906).

CHEBYSHEV, P. L.

[16] On mean values. J. Math. Pures Appl. (2) **12**, 177–184 (1867).

---

† *Translator's note.* Titles in Russian are translated into English and are starred by an asterisk. If a reference is available both in Russian and in English, French, or German, usually only the latter is cited here; in some cases both the (Russian) original and the translation are given. However, no research has been made into possible existing translations of Russian works. The abbreviations used here follow roughly the system adopted by the *Mathematical Reviews*, as does the transliteration of Russian names. Several less familiar Russian periodicals are transliterated *ad hoc*. Note that the numbering of the references is different from that in the original.

[17] Sur deux théorèmes relatifs aux probabilités. Acta Math. **14**, 305–315 (1890).
[18] Sur le développement des fonctions à une seule variable. Bull. de l'Acad. Imp. des Sci. de St. Petersbourg (3) **1**, 193–200 (1860).

CRAMÉR, H.
[19] On the composition of elementary errors. Skand. Aktuarietidskr. 13–74; 141–180 (1928).
[20] Ueber eine Eigenschaft der normalen Verteilungsfunktion. Math. Z. **41**, 405–414 (1936).
[21] Random variables and probability distributions. Cambridge Tracts in Mathematics no. 36, Cambridge, 1937.
[22] Mathematical methods of statistics, Princeton, 1946.

DOEBLIN, W.
[23] Sur les sommes d'un grand nombre de variables aléatoires indépendantes. Bull. Sci. Math. **63**, no. 1, 23–32, 35–64 (1939).
[24] Sur l'ensemble de puissances d'une loi de probabilité. Studia Math. **9**, 71–96 (1940).

EDGEWORTH, F. Y.
[25] The law of error. Proc. Cambridge Philos. Soc. **20**, 36–65 (1905).

ESSEEN, C. G.
[26] Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. Acta Math. **77**, 1–125 (1945).

FELLER, W.
[27] Ueber den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. Math. Z. **40**, 521–559 (1935); **42**, 301–312 (1937).
[28] Ueber das Gesetz der grossen Zahlen. Acta Univ. Szeged. Sect. Sci. Math. **8**, 191–201 (1937).
[29] Neuer Beweis für die Kolmogoroff-P. Lévysche Characterisierung der unbeschränkt teilbaren Verteilungsfunktionen. Bull. Intern. Acad. Yougoslave. Cl. Sci. Math. Nat. **32**, 1–8 (1939).

DE FINETTI, B.
[30] Sulla funzione a incremento aleatorio. Atti Acad. Naz. Lincei. Rend. Cl. Sci. Fis. Mat. Nat. (6) **10**, 163–168, 325–329, 548–553 (1929).
[31] Le funzione caratteristische di legge istantanea. Ibid. (6) **12**, 278–282 (1930).

GLIVENKO, V. I.
*[32] The Stieltjes integral. GONTI, Moscow-Leningrad, 1936.
*[33] A course in the theory of probability, GONTI, Moscow-Leningrad, 1939.
[34] Sur les valeurs probables de fonctions. Atti Acad. Naz. Lincei. Rend. Cl. Sci. Fiz. Mat. Nat. (6) **8**, 480–483 (1928).

GNEDENKO, B. V.
[35] Sur une propriété caractéristique des lois infiniment divisibles. Bull. Univ. d'Etat Moscou. Sér. Internat. Sect. A. **1**, no. 5, 9–15 (1937).
[36] Über die Konvergenz der Verteilungsgesetze von Summen voneinander unabhängiger Summanden. C. R. (Doklady) Acad. Sci. URSS (N.S.) **18**, 231–234 (1938).
*[37] On the theory of limit theorems for sums of independent random variables. Izvestiya Akad. Nauk SSSR, Ser. Mat. 181–232, 643–647 (1939).
*[38] On the theory of domains of attraction of stable laws. Ucheny̆e Zapiski Moskov. Gos. Univ. Matematika **30**, 61–81 (1939).
*[39] Some theorems on the powers of distribution functions. Ucheny̆e Zapiski Moskov. Gos. Univ. Matematika **45**, 61–72 (1940).
[40] Sur la distribution limite du terme maximum d'une série aléatoire. Ann. Math. **44**, 423–453 (1943).
[41] Limit theorems for sums of independent random variables. Uspekhi Mat. Nauk

**10,** 115–165 (1944). English translation: Translation no. **45,** American Mathematical Society, New York.

*[42] On a local limit theorem in the theory of probability. Uspekhi Mat. Nauk (N.S.) **3,** no. 3, 187–194 (1948).

*[43] On a local limit theorem in case of infinite variance. Trudy Mat. Inst. Akad. Nauk. Ukrain. SSR **12,** 22–30 (1949).

*[44] On some properties of limit distributions for normalized sums. Ukrain. Mat. Zhurnal, no. 1, 3–8 (1949).

*[45] On a local theorem for stable limit distributions. Ukrain. Mat. Zhurnal, no. 4, 3–15 (1949).

GNEDENKO B. V. and GROSHEV, A. V.

*[46] On the convergence of distribution laws of normalized sums of independent random variables. Rec. Math. [Mat. Sbornik] N.S. **6** (**48**), 521–541 (1939).

GROSHEV, A. V.

*[47] The domain of attraction of the Poisson law. Izvestiya Akad. Nauk SSSR. Ser. Mat. **5,** 165–172 (1941).

HARDY, G., LITTLEWOOD, J. E. and PÓLYA, G.

[48] Inequalities. Cambridge, 1934.

KHINTCHINE, A. YA.

*[49] Fundamental laws in the theory of probability. 1926; GTTI, Moscow-Leningrad, 1932.

[50] Begründung der Normalkorrelation nach der Lindebergschen Methode. Izvestiya Assoc. Inst. Moskov. Univ. **1–2,** 37–45 (1928).

[51] Sur la loi des grands nombres. C. R. Acad. Sci. Paris, **188,** 477–479 (1929).

[52] Sul dominio di attrazione della legge di Gauss. Giorn. Ist. Ital. Attuari **6,** 371–393 (1935).

[53] Asymptotische Gesetze der Wahrscheinlichkeitsrechnung. Ergebnisse der Mathematik, II. 4, Springer, Berlin, 1933.

[54] Su una legge dei grandi numeri generalizzata. Giorn. Ist. Ital. Attuari **7,** 365–377 (1936).

[55] Ueber Klassenkonvergenz von Verteilungsgesetzen. Izv. Nauchno-Issled. Inst. Mat. Meh. Tomsk. Gos. Univ. **1,** 258–261 (1937).

[56] Déduction nouvelle d'une formule de M. Paul Lévy. Bull. Univ. d'Etat Moskou. Sér. Internat. Sect. A. **1** no. 1, 1–5 (1937).

[57] Contribution à l'arithmétique des lois de distribution. Ibid. **1,** no. 1, 6–17 (1937).

[58] Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze. Rec. Math. [Mat. Sbornik] N.S. **2** (**44**), 79–120 (1937).

*[59] Limit theorems for sums of independent random variables. GONTI, Moscow-Leningrad, 1938.

*[60] On unimodal distributions. Izv. Nauchno-Issled. Inst. Mat. Meh. Tomsk. Gos. Univ. **2,** no. 2, 1–7 (1938).

[61] Mathematical foundations of statistical mechanics. GTTI, Moscow-Leningrad, 1938. English translation by G. Gamow, Dover, New York, 1949.

KHINTCHINE A. YA. and LÉVY, P.

[62] Sur les lois stables. C. R. Acad. Sci. Paris **202,** 374–376 (1936).

KOLMOGOROV, A. N.

[63] Ueber die Summen durch den Zufall bestimmter unabhängiger Grössen. Math. Ann. **99,** 309–319 (1928); **102,** 484–488 (1929).

[64] Sulla forma generale di un processo stocastico omogeneo. Atti Acad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat. (6) **15,** 805–808, 866–869 (1932).

[65] Grundbegriffe der Wahrscheinlichkeitsrechnung. Ergebnisse der Mathematik II. 4, Springer, Berlin, 1933.

[66] Eine Verallgemeinerung des Laplace-Liapounoffschen Satzes. Izvestiya Akad. Nauk SSSR, Ser. Mat., 956–962 (1931).
[67] Ueber die Grenzwertsätze der Wahrscheinlichkeitsrechnung. Izvestiya Akad. Nauk SSSR, Ser. Fiz-Mat., 363–372 (1933).
[68] Ueber die analytischen Methoden in der Wahrscheinlichkeitsrechnung. Math. Ann. **104**, 415–458 (1931).
*[69] Theory of measure and Lebesgue integral. (In press.)
KOZULYAEV, P. A.
[70] Asymptotic analysis of a fundamental formula in the theory of probability. Uchenye Zapiski Moskov. Gos. Univ. Matematika **15**, 179–182 (1939).
LAPIN, A. I.
[71] On some properties of stable laws. Dissertation, 1947.
LAPLACE, R.
[72] Théorie analytique des probabilités. Paris, 1812, 1814, 1820.
LÉVY, P.
[73] Calcul des probabilités. Gauthier-Villars, Paris, 1925.
[74] Sur les intégrales dont les éléments sont des variables aléatoires indépendantes. Annali R. Scuola Norm. Sup. Pisa (2) **3**, 337–366 (1934); **4**, 217–218 (1935).
[75] Propriétés asymptotiques des sommes de variables aléatoires indépendantes ou enchaînées. J. Math. Pures Appl. (9) **14**, 347–402 (1935).
[76] Théorie de l'addition des variables aléatoires. Gauthier-Villars, Paris, 1937.
LINDEBERG, Y. W.
[77] Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. Math. Z. **15**, 211–225 (1922).
LINNIK, YU. V.
[78] On the accuracy of approximation to the Gaussian distribution by sums of independent random variables. Izvestiya Akad. Nauk SSSR, Ser. Mat. **11**, 111–138 (1947).
LYAPUNOV, A. M.
[79] Sur une proposition de la théorie des probabilités. Bull. de l'Acad. Imp. des Sci. de St. Petersbourg (5) **13**, no. 4, 359–386 (1900).
[80] Nouvelle forme du théorème sur la limite de probabilités. Mem. Acad. Sci. St. Petersbourg (8) **12**, no. 5, 1–24 (1901).
MARCINKIEWICZ, J.
[81] Sur les fonctions indépendantes. II. Fund. Math. **30**, 349–364 (1938).
MARKOV, A. A.
*[82] The law of large numbers and the method of least squares. Izv. Fiz.-Mat. Obshchestva Kazan. Univ. (2) **8**, 110–128 (1899).
[83] Sur les racines de l'équation $e^{x^2}\dfrac{d^m e^{-x^2}}{dx^m} = 0$. Bull. de l'Acad. Imp. des Sci. de St. Petersbourg (5) **9**, 435–446 (1898).
*[84] Calculus of probability. 1900, 1908, 1913, 1924. German translation of 1908 ed.: Wahrscheinlichkeitsrechnung, Teubner, Leipzig-Berlin, 1912.
MEĬZLER, D. G., PARASYUK, O. S., RVACHËVA, E. L.
*[85] On a multidimensional local limit theorem in the theory of probability. Doklady Akad. Nauk SSSR **60**, 1127–1128 (1948); Ukrain. Mat. Zhurnal, no. 1, 9–20 (1949).
DE MOIVRE, A.
[86] Miscellanea Analytica Supplementum. London, 1730.
POISSON, S. D.
[87] Recherches sur la probabilité des jugements. Paris, 1837.
RAĬKOV, D. A.
*[88] On the decomposition of the Gaussian and Poisson Laws. Izvestiya Akad. Nauk. SSSR, Ser. Mat. 91–124 (1938).

*[89] On a connection between the central limit theorem in the theory of probability and the law of large numbers. Izvestiya Akad. Nauk SSSR, Ser. Mat., 323–338 (1938).

ROMANOVSKIĬ, V. I.

[90] Sur une extension du théorème de A. Liapounoff sur la limite de probabilité. Izvestiya Akad. Nauk SSSR, Ser. Mat., 209–225 (1929).

SLUTSKY, E. E.

[91] Ueber stochastische Asymptoten und Grenzwerte. Metron **5**, 1–90 (1925).

TITCHMARSH, J.

[92] Theory of Fourier integrals. Oxford, 1937.

ZHURAVSKIĬ, A. M.

*[93] On a limit theorem in the calculus of probability, Trudy Fiz.-Mat. Inst. imeni Steklov. **4**, 9–36 (1933).

# INDEX


Accompanying laws, 98
Additive characteristic, 44
Asymptotic expansion, 10
  for continuous distributions, 220
  for densities, 228
  failure of, 222
  for lattice distributions, 241
  for nonlattice distributions, 210
Asymptotically constant summands, 95
Axiomatics, 20

Bernoulli distribution, 217
Berry-Esseen theorem, 201
Borel closure, 16
Borel field, 16

Canonical decomposition of set functions, 30
Canonical representation, of infinitely divisible laws, 76
  of laws of the class $L$, 149
  of stable laws, 164
  uniqueness of, 76, 80
Cauchy's law, 72
  local limit theorem for, 236
Central limit theorem (*see also* Normal distribution), 3, 4
Characteristic exponent of a stable law, 171
Characteristic function(s), 44
  convergence of a sequence of, 53
  derivatives of, 63
  inequalities on, 53, 55
  inversion formula, 48
  logarithm of, 64
  uniqueness of, 50
Chebyshev-Hermite polynomials, 191
Chebyshev's expansion, 191
Chebyshev's two problems, 4
Chi-square ($\chi^2$) distribution, 72
Class $L$, 145
  canonical representation, 149
  criterion for, 147
  example with finite variance, 152
Complete inverse image, 17
Composition, 28
Condition ($\omega$), 232
Conditional mathematical expectation, 21, 248, 250
Conditional probability, 21, 248, 250
Conditionally compact set of distributions, 38
Convergence, of densities, 222
  of distributions and measures (*see* Weak convergence)
  failure of, 223
  in probability, 105
  of series, 137
  of sums of independent summands (*see also* Limit distributions and Domain of attraction), most general form, 98–101, 112, 116–124
    for normalized sums, 152–157
  of variances, 97
  with probability one, 137
Cramér's condition ($C$), 209
Cramér's theorem, on asymptotic expansion, 220
  on normal distribution, 51

DeMoivre-Laplace theorem, 5
Density, 23
Distance (of Lévy), 33
Distribution function (*see also* Normal distribution, Probability distribution), 14, 24
Domain, of attraction, 172
  of normal attraction, 181
  of a normal law, 172
  of partial attraction, 184
  of a stable law, 175
  theorems on, 189–190

Elementary event, 20
Exponential distribution, 9

Feller's form of the central limit theorem, 130
Field of sets, 16
  Borel, 16
  unit of, 16

Gnedenko's theorem, on convergence of infinitely divisible laws, 87
  on convergence of sums, 112

Identically distributed random variables, 162
Improper distribution, 24
  theorems on, 56
Incomplete gamma function, 9
Independence, 26, 250
Infinitely divisible distributions, 71
  canonical representation, 76
  convergence of a sequence of, 87
  examples, 71–76
  examples with striking properties, 81-83
  generation by Poisson laws of, 74
  as limit distributions in partial attraction, 184
  as limit distributions of sums of infinitesimal summands, 115
Infinitely divisible random variable, 71
Infinitesimal summands, 95
  criterion for, 96
  uniformly, 126
Inversion formula, 48

Khintchine's theorem, on characterization of infinitely divisible laws, 115–116
  on partial attraction, 184
Kolmogorov-Khintchine criterion for convergence of series, 137
Kolmogorov's formula, 85

Laplace's theorem, 2
Lattice distribution, 58, 212, 231
  (maximum) span of, 58, 60, 232
Law of large numbers (*see also* Relative stability), 3, 4, 105, 133–139
  Khintchine's form of, 138
  Kolmogorov-Feller form of, 135
Lebesgue integral, 19
Lévy-Khintchine formula, for infinitely divisible laws, 70
  for stable laws, 62
Lévy's formula, 84
Lindeberg's condition, 5
Lindeberg-Feller theorem, 103
Lyapunov's condition, 5, 103
Lyapunov's theorem, 201
Local limit theorems, 231

Markov's condition, 4
Mathematical expectation, 14
Measure, 17
  carrier of, 17
  complete, 250
  generated, 17
  normalization of, 20
  perfect, 18
  in product space, 27
Median, 95
Measurable, function, 19
  mapping, 17
  set, 19
Method of moments, 4
Metric space, of characteristic functions, 52
  of distributions, 37
Moment, 62
  absolute, central, absolute central, 62

Normal distribution, 3
  convergence to, 102, 126–132, 143, 172, 181
  convergence to density of, 224, 228
  as infinitely divisible, 71
  local limit theorems for, 233, 241
  special role of, 126
Normalized sum, 145

Pearson curves, 72
  as infinitely divisible, 86

Poisson's law, 47
convergence to, 104, 132
generating infinitely divisible laws, 74
as infinitely divisible, 72, 247
special position in classical probability, 8, 145
Poisson's limit theorem for rare events, 3, 8
generalized, 8
Possible value, 23
Principal branch, 64
Probability, 20
density, 23
distribution, 22
continuous, discrete, 23
joint, 22
$n$-dimensional, 23
proper, improper, 24

Raikov's theorem, on Poisson's law, 51
on relative stability, 143
Random, event, 21
function, 68
variable, 21
critique of, 13
existence of, 246
vector, 22
walk, 6
Relative stability, 139
Remainder term, 196
estimation of, 201
extremal case, 217
lattice case, 212
nonlattice case, 208
Riemann zeta-function, 75

Semi-invariant, 65
Stable laws, 162
canonical representation of, 164
convergence to, 175, 181
convergence to density of, 227
examples of, 171
local limit theorem for, 236
properties of, 182–183
Stable sequence, 105
Stable type, 162
Stieltjes integral, 29
discussion of, 15
Stieltjes sums, 26
Stochastic process with independent increments, 127
Stochastically (strongly) continuous process, 128
Symmetrical distribution function, 51

Theory of errors, 3
Type of distribution functions, 40
improper, proper, 40

Unimodal distribution function, 157, 252
criteria for, 157, 160
examples of, 253
vertex of, 157
Uniqueness theorem, for canonical representation, 76, 80
for characteristic function, 50
Unitary law, 7
convergence to (*see* Law of large numbers)
theorems on, 57–58
Universal law (of Doeblin), 189

Variance, 44
Variation, negative, 30
positive, 30
total, 30

Weak convergence, characterizations of, 33
of distributions, 32
of measures, 39
in terms of characteristic functions, 53